Genomics for Optimal Aphid Biocontrol

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Mark Andrew Whitehead

September 2019

UNIVERSITY OF LIVERPOOL

James Hutton Limited

Acknowledgments

I am forever grateful to everyone who has helped me get to this position, and would be impossible to thank them all individually.

Firstly, I would like to thank the group at Liverpool: Joe, Sam, Lauren (especially for putting up with me for the last five years now!), Alex, Dave, Fran and Stef, as well those at the JHI: Jenny, Daniel, Carolyn, Marta and Desire'. Many thanks to the CGR and the bioinformatics crew (Sam, Richard and Matthew) for every time I've pestered them for help. I would also thank Adam, Matt, Joel and Jamie who all, at one point or another, helped make the flat a home over the past four years. I'm especially thankful to my supervisors, Alistair and Ali, for their constant guidance and knowledge throughout the course of my PhD, as well as the opportunity to do this project.

I want to thank my Mum and Dad for their unwavering support and encouragement. Finally, I want to thank Jen, for being my rock throughout my PhD and whom I would be totally lost without.

Preface

Many aphid species cause the reduction of crop yield and monetary loss through feeding and pathogen vectoring. Current pest management strategies for aphids require a range of control methods. However, increased resistance to these methods in certain populations of aphids could impact successful control of these pests. This thesis focuses specifically on the Potato aphid, *Macrosiphum euphorbiae*, and its innate resistance to the parasitic wasp, *Aphidius ervi*, a common natural enemy of multiple aphid species. I first discuss population dynamics of *M. euphorbiae* within the UK and its existence as a handful of strictly asexual genotypes. The distribution is skewed towards genotypes potentially linked to increased host range, while the genotype with low susceptibility to *A. ervi* (genotype 1) has lower frequency in comparison. A cost associated with *A. ervi* parasitism-resistance in potato aphids has yet to be demonstrated, while also exhibiting no detrimental effects on fitness in a response to *A. ervi* challenge as well as potentially having increased tolerance to sub-zero temperatures. To further study innate parasitoid resistance, I generated a draft genome of the genotype 1 Potato aphid (and its obligate symbiont *Buchnera aphidicola*) using a mixture of long-read sequencing and chromosome scaffolding techniques. Comparative genomics between this aphid genotype with other genotypes of Potato aphid and other aphid species provides a list of immune functioning genes that could play a role in immunity, either through sequence variation or copy number aberrations. However, it is difficult to highlight any specific pathway as the cause of parasitism-resistance to *A. ervi*. Metabolome comparison between Potato aphid genotypes 1 and 2 highlights variance between genotypes and in response to parasitoid challenge, and suggests the mounting of an immune response through the increased conversion of trehalose to glucose, previously observed in cellular immunity. The data generated over the course of this project will provide useful information for integrated pest management of Potato aphids, as well as providing more resources for further comparative genomics studies within Aphididae.

Contents

1. General Introduction 1
1.1 The aphid pest 1
1.2 A note on the "usefulness" of aphids 3
1.3 Obligate asexual reproduction in aphids and its implications 4
1.4 The importance of Integrated Pest Management (IPM) 6
1.5 Bio-control methods for aphids and the parasitoid *Aphidius ervi* 7
1.6 Role of the aphid obligate symbiont 9
1.7 Resistance to parasitism and other useful traits of symbiosis 10
1.8 The role of monitoring in IPM strategies 12
1.9 The Potato aphid *Macrosiphum euphorbiae* 13
1.10 Current IPM strategies regarding *M. euphorbiae* 16
1.11 Innate resistance to parasitism in other insects 17
1.12 Aims and objectives 19

2. General methods 20
2.1 Insect culture and maintenance 20
2.2 Genomic DNA extraction and preparation of short read Illumina libraries 20
2.3 *M. euphorbiae* challenge with the parasitoid *Aphidius ervi* 21
2.4 Data sources used in analysis 21

3. Population structure insights into *M. euphorbiae* 22
3.1 Introduction 23
3.1.1 Phenotypic and genotypic variance within aphid species 23
3.1.2 Genotypic variance and impact on general aphid fitness 24
3.1.3 Advantageous survival traits may exist in a cost-benefit system 25
3.1.4 Chapter aims 26
3.2 Methods 28
3.2.1 Sampling strategy for *M. euphorbiae* 28
3.2.2 Insect culture and maintenance 30
3.2.3 DNA extraction for genotyping and symbiont infection 30
3.2.4 Genotyping collected aphid lines 31
3.2.5 Experimental aphid lines 32
3.2.6 Confirming parasitoid susceptibility in 2016 clonal lines 32
3.2.7 Assessing aphid response to parasitoid attack 33
3.2.8 Aphid survival in response to cold temperatures 34
3.2.9 Statistical analyses 34
3.3 Results 36
3.3.1 Genotype distributions of *M. euphorbiae* 36
3.3.2 STRUCTURE analysis supports strict clonality 42
3.3.3 Preliminary resistance assays confirm high levels of resistance in samples of genotype 1 44
3.3.4 Parasitoid attack does not affect aphid development and fecundity in genotype 1 46
3.4 Discussion 52
3.4.1 Genotype 1 advantageous traits do not confer high abundance in *M. euphorbiae* populations 52
3.4.2 Parasitoid resistant clones show no detrimental effects in response to attack 53
3.4.3 Genotype 1 shows greater cold tolerance than genotypes 2 and 3 54
3.4.4 Conclusion 55

3.5 Supplementary tables **56**
3.6 Supplementary figures **63**
3.7 Other electronic resources **64**

4. Genome assembly of parasitoid resistant *Macrosiphum euphorbiae* **65**
4.1 Introduction **66**
4.1.1 The rise of insect genome sequencing 66
4.1.2 Gene annotation of current aphid genomes 69
4.1.3 Aims and objectives 70
4.2 Methods **72**
4.2.1 *M. euphorbiae* culture conditions for sequencing material 72
4.2.2 Genomic DNA extraction and preparation of short read Illumina libraries 72
4.2.3 High molecular weight DNA preparation and Long-read library sequencing 72
4.2.4 RNA extraction for transcriptome sequencing 74
4.2.5 Genome assembly of *M. euphorbiae* 75
4.2.6 Genome assembly polishing using arrow 75
4.2.7 Assessment of genome contiguity and gene completeness 75
4.2.8 Removing diploid sequence through Haplomerger2 76
4.2.9 Scaffolding and contig sorting into chromosomal groups 76
4.2.10 Visualising symbionts and assembly contamination with Blobtools 77
4.2.11 Data sources for comparative analyses 77
4.2.12 Genome synteny between other aphids 77
4.2.13 Repeatmasking and repeat element annotation 78
4.2.14 Transcriptome assembly of MinION and Illumina reads 78
4.2.15 Draft genome annotation using Maker2 78
4.2.16 Draft genome annotation using BRAKER 79
4.2.17 Filtering predicted gene sets on expression 79
4.2.18 Identification of immune function genes in *M. euphorbiae* 80
4.2.19 Genome assembly for the Potato aphid obligate symbiont, *Buchnera aphidicola* 80
4.2.20 Gene prediction and pseudogene identification for *B. aphidicola* 81
4.2.21 Sequence alignment of individual *Buchnera* genes 82
4.2.22 Phylogenetic analysis of *M. euphorbiae* and *B. aphidicola* in the aphid clonal lines 82
4.3 Results and Discussion **83**
4.3.1 Contig assembly and scaffolding of the *M. euphorbiae* genome 83
4.3.2 Genome scaffolding 87
4.3.3 Repetitive elements of *M. euphorbiae* 92
4.3.4 Transcriptome assembly and gene prediction with MinION and Illumina transcriptomic data 94
4.3.5 Phylogenetic analysis of *M. euphorbiae* 98
4.3.6 Shared immune function proteins between *M. euphorbiae* and *A. pisum* 99
4.3.7 Genome assemblies for *B. aphidicola* of *M. euphorbiae* 102
4.4 Conclusion **109**
4.5 Supplementary tables **110**
4.6 Supplementary Figures **111**
4.7 Other electronic resources **119**

5. A multi-omic approach to assess clonal differences in genotypes of *M. euphorbiae* **120**
5.1 Introduction **121**
5.1.1 Effect of copy number variation on aphid phenotype 121

5.1.2 Gene copy numbers with regards to the aphid immune system 122
5.1.3 The potential of Metabolomics 124
5.1.4 Specialised parasitoid mechanisms for survival within a host 126
5.1.5 Aims and objectives 127
5.2 Methods 129
5.2.1 Genomic DNA extraction and preparation of short read Illumina libraries 129
5.2.2 Read mapping to the M. euphorbiae reference genome 129
5.2.3 Variant calling between aphid genotypes 130
5.2.4 Assessing genetic relationship between Potato aphid genotypes 131
5.2.5 Generating genotype 1 specific gene variants 132
5.2.6 Variant calling between genotypes of the *Buchnera* symbiont 132
5.2.7 Assessing presence of copy number variation between parasitism-resistant and -susceptible Potato aphid genotypes 133
5.2.8 dN/dS analysis of genotype 1 paralogues genes single copy orthologues to the Pea aphid 133
5.2.9 Parasitoid assays for metabolomics analyses 135
5.2.10 Sample processing for 1H NMR 135
5.2.11 Metabolite identification 136
5.2.12 Metabolite normalisation and scaling 137
5.2.13 Statistical analyses to identify significantly different metabolite peaks 137
5.3 Results 138
5.3.1. Sequencing results 138
5.3.2 Clonality in *M. euphorbiae* genotypes is observed at both the level of aphid host and symbiont 138
5.3.3 dN/dS highlight immune function genes of interest 145
5.3.4 CNV analysis highlights increases in copy number prominent genotype 1 149
5.3.5 Identified metabolites of *M. euphorbiae* 152
5.3.6 Variance between metabolomes of genotypes 1 and 2 154
5.3.7 Metabolomic profiles of *M. euphorbiae* challenged with *A. ervi* 156
5.4 Discussion 163
5.4.1 Whole genome sequencing (WGS) helps elucidate Potato aphid genotypes. 163
5.4.2 Copy number aberrations and their link to multiple Potato aphid phenotypes 163
5.4.3 Minor variations between Potato aphid metabolomes 165
5.4.4 Conclusion 166
5.5 Supplementary figures 167

6. General discussion 169
6.1 Summary of the work 169
6.1.1 Rationale for studying *M. euphorbiae* 169
6.1.2 Further insights into UK populations of *M. euphorbiae* 170
6.1.3 Contribution of the *M. euphorbiae* genome to the field 170
6.1.4 Multiple candidate causes of *A. ervi* resistance 171
6.1.5 A possible role for *Buchnera* in genotypic variation 172
6.2 Future work 173
6.2.1 Genomic and metabolomic resources for *M. euphorbiae* 173
6.2.2 Further study into parasitoid resistance and other resistance traits in *M. euphorbiae* 174
6.2.3 Breeding programmes for overcoming resistance 175
6.3 General Conclusions 176

References 177

1. General Introduction

1.1 The aphid pest

Protection of crops from pests involves the reduction of yield associated to a range of detrimental factors. In this instance, “pest” mainly refer to weeds, insects and pathogens they can spread. While weeds are arguably more of a threat, insect pests and viral pathogens both pose a considerable issue for crop growth, predicted to lead to approximately 18% and 16% losses in reduction respectively (Oerke, 2004; Oerke, 2006).

The family Aphididae consists of approximately 5,000 individually described species, 455 of which are found on crops (Blackman & Eastop, 2000). Aphids are one of the more successful and devastating crop pests, and a significant vector for plant pathogens, with around 100 of which considered to cause major economic impact (Van Emden & Harrington, 2017). What makes these insects so successful relies on aspects related to their life-cycle and their mode of reproduction. All aphid species undergo cyclical parthenogenesis (CP), producing both asexual and sexual morphs depending on external factors and environmental conditions (Dixon, 1973). In aphids, summer morphs are asexual and viviparous, giving birth to live young. Nymphs develop within aphid ovarioles (figure 1), with these nymphs also harbouring their own developing embryos. Therefore, adult females will at some stage carry not only their own offspring, but their developing granddaughters as well. The success of aphids and their ability to reproduce exponentially is attributed to viviparity and these telescopic generations (Dixon, 1973). Increased density on a host plant of asexual morphs can promote morphological divergence in developing nymphs, where increased contact between nymphs results in development into winged morphs (alates) for dispersal (Lees, 1967). Production of winged morphs is also a useful adaptive trait for colony survival in the presence of increased natural enemy abundance and activity; when host-plant quality is poor and nutrient provision is reduced; and for movement to winter hosts in the sexual phase of the life cycle (Müller *et al.,* 2001;

Mehrparvar *et al.,* 2013). As summer crops (secondary host plants) are not cultivated year-long, a winter host plant (primary host plant) is required to sustain aphids until the following year, which coincides with another morphological change and the switch to gynoparous females that in turn can generate males and oviparous (egg-laying) females (Dixon, 1973). Production of sexual morphs is induced by multiple environmental cues, but mainly through the shortening of day-length and reductions in temperature associated with autumn (Dixon, 1973). Sexual reproduction between males and oviparous females generates cold-hardy eggs to survive sub-zero temperatures (Strathdee *et al.,* 1995). The life-cycle culminates in egg hatching and development into a fundatrix specially adapted to have high levels of fecundity, producing asexual summer morphs previously described (Wellings *et al.,* 1980; Van Emden & Harrington, 2017).

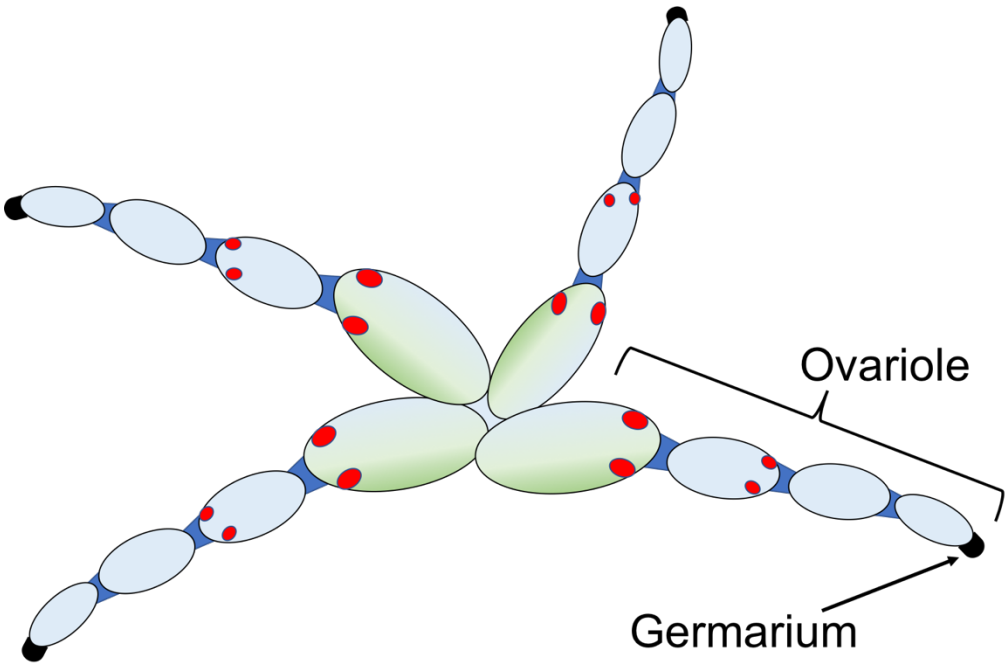


Figure 1. Example ovariole cluster within the aphid ovary during parthenogenesis. Aphid oogenesis begins at the germarium where eggs are ovulated from, and are radially dispersed within the ovary. Each ovary has approximately 4-6 of these ovarioles, and two ovaries per aphid. Aphids develop in series until they are fully developed nymphs, then born via the aphid oviduct (Dixon, 1973; pictured adapted from Miura *et al.,* 2003).

The ability of aphids to give birth to live young, as well as disperse as winged forms can result in them being highly successful pests. For some aphid species, survival probability is increased by the capacity to infest and reproduce upon multiple host plants. Polyphagy also increases the likelihood that an aphid species will become a major crop pest through the spread of multiple plant viruses (Ng & Perry, 2004), with virus damage one of the major components of crop damage and reduction in yield.

The acquisition of viruses in aphids is linked to the behaviour exhibited during feeding, and can be loosely grouped into persistent viruses, which remain and replicate in aphid tissues for weeks or months at a time, and non-persistent viruses that cling to the interior of the stylet or salivary glands, remaining on the aphid for only hours or days (Dietzgen *et al.,* 2016). The act of probing by the aphid stylet to assess nutrient provision from a potential host plant provides ample time for any non-persistent virus to attach within the aphid stylet and includes a range of Potyviruses and various mosaic viruses, as reviewed in Pirone and Harris (1977).

1.2 A note on the "usefulness" of aphids

While some aphid species are considered a major agricultural pest, they also have unique characteristics that make them good model organisms for studying a wide range of genetic and phenotypic traits. The clonal nature of reproduction of the summer morphs means they are (generally) easy to maintain and culture, being able to generate large quantities of study material due to their exponential growth in numbers. Clonal lineages are also helpful in the study of phenotypic traits linked to different genotypes and the evolution of traits that provide a selective advantage, for example through the interactions of aphids with insect natural enemies and aphid pathogens (Ferrari *et al.,* 2001; Li *et al.,* 2002). Aphids have provided a useful model for studying the dynamics of host-symbiont interactions with the primary symbiont *Buchnera* (Douglas *et al.,* 2010) and many facultative or secondary symbionts (Darby & Douglas 2003; Ferrari *et al.,* 2004), as

well as the ecological consequences of symbiont infection (Darby *et al.,* 2003; Ferrari *et al.,* 2010).

Genome and transcriptome sequencing of aphids also permits the study of genome evolution across the Aphididae family. For example, transcriptomics has previously been used to assess variation in aphid effector proteins between aphid species (Thorpe *et al.,* 2016), where effector proteins can have an impact on success of an aphid on a host as well as its host plant range. Currently, the web-portal AphidBase hosts seven aphid genome projects (Legeai *et al.,* 2010), followed by a further genome project for the Corn Leaf Aphid (*Rhopalosiphum maidis*) on NCBI (Chen *et al.,* 2019). While the study of aphids at the genome level is useful for uncovering the genetic basis behind certain phenotypes such as insecticide resistance or host range, the clonal nature of these organisms circumvents issues normally associated genome sequencing, such as the use of multiple individuals and the heterozygosity this may introduce that ultimately hinders genome assembly.

1.3 Obligate asexual reproduction in aphids and its implications

Aphid populations can sometimes be comprised of groups with alternate forms of reproduction. Where many will exist through cyclical parthenogenesis as previously described, others may persist asexually over-winter without the production of eggs (Delmotte *et al.,* 2001; Papura *et al.,* 2003). Factors usually implicated in inducing the switch to gynoparae such as day-length and temperature fail to trigger sexual morph production, and aphids maintain a strictly asexual method of reproduction (also known as obligate parthenogenesis, or OP) (Dixon, 1973). Shortening of days usually affects the production of juvenile hormones and other signalling compounds, causing changes in aphid ovarioles which dictate that cells should undergo meiosis and generate haploid eggs (Simon *et al.,* 2010). Sex determination and generation of male/female offspring operates via an XX/XO system, where only females are diploid for the X chromosome (XX) and males are haploid (XO), inheriting a random X chromosome from the mother

(Caillaud *et al.,* 2002; Jaquiéry *et al.,* 2011). Why this switch is ineffective in obligate asexual aphids is still debated, although candidate genes have been identified through the study of cyclical and obligate parthenogens of the Cotton aphid, where expression of genes in phototransduction (sensory perception of light) and embryogenesis/oogenesis, for example, are differentially expressed (Liu *et al.,* 2014). However, it is likely that genetic changes that underpin the change to strict asexuality can occur through multiple routes (Delmotte *et al.,* 2001). The pressures that lead to maintenance of asexuality are unclear, considering a lack of sexual reproduction would remove the opportunity to remove deleterious mutations that may occur and could slow down the ability to select those who adapt to a changing environment. It has been suggested that asexuality could be maintained if environmental conditions (e.g. daylength, temperature) do not vary enough between changing seasons to induce the switch to sexual forms, leaving these genes open to deleterious mutations (Simon *et al.,* 2002). Some asexual lines of aphids will also maintain the ability to generate sexual males, where these sexual males proceed to mate with sexual females and pass down genes for strict asexuality. This is also known as contagious asexuality (Simon *et al.,* 1999; Simon *et al.,* 2002).

Asexual lineages within aphid populations can play an interesting role with regards to pest control and integrated pest management (IPM) (see section 1.5). Specifically, the idea that specific asexual lineages can be under strong positive selection due to expressing a trait or traits that greatly enhance survival and offspring production, leading to the predominance of “superclones”. For example, recent outbreaks of *Rhopalosiphum padi* (Linnaeus) in Chile mainly consisted of a single asexual genotype with high fecundity (Rubio-Meléndez *et al.,* 2019). Another example relates to genotype distributions of the Australian cotton aphid, *Aphis gossypii* (Glover), dominated by a superclone that presents primicarb resistance (Chen *et al.,* 2012). The ability to detect these phenotypes again highlights the importance of monitoring and how best to inform IPM going forward.

1.4 The importance of Integrated Pest Management (IPM)

Phloem sap-feeding insects pose a major threat to agriculture and farming industries, causing direct damage as a result of feeding or more importantly indirect damage via the spread of harmful viral pathogens. If left unchecked, monetary losses associated with decreased crop yields could easily fall into the billions (Scholthof *et al.,* 2011). The role of IPM is to outline the best strategy for control of a given pest, ideally through a combination of different pest management and control methods while also accommodating a range of beneficiaries and stakeholders (Korgan, 1998). The main principle that governs IPM is that chemical and biological methods of control are complementary to one another, while further promoting and enhancing naturally occurring protection (Smith *et al.,* 1976). However, this holistic approach to designing IPM strategies is often more complex, also considering other factors such as economic pressures of the area in question, local agricultural policy and even potential effects of local climate and weather (Norton & Mumford, 1993). IPM itself should be a dynamic process, with constant monitoring and evaluation of pests and using this to modify the outlined IPM strategy. A simple example could be fluctuations in climate and its effect on crop growth. This could have multiple impacts on crop pests and their natural enemies and therefore inform choices on chemical pesticide usage. Being able to monitor these situations can also better inform predictive models of pest population dynamics, also a useful tool for determining IPM strategies (Jervis & Kidd, 1996).

Chemical pesticides such as insecticides are some of the more commonly used methods of pest control. While often very effective and more immediate than potentially slower acting methods of biocontrol, their benefits are somewhat-outweighed by their harmful interaction with the surrounding environment (Willis & McDowell, 1982). A common and more recent problem that has arisen through the use of chemical pesticides is their effect on non-target organisms, such as the use of neonicotinoids and the detrimental effect on honeybees (Johnson *et al.,* 2010), the knock-on effect being a reduction in pollen dispersal. Chemical

pesticides may interfere with natural enemies of target arthropod pests. The efficiency of biocontrol, which can often rely on the use of these predatory species, could be dramatically reduced as a result of insecticide usage (Longley, 1999; Joseph *et al.,* 2010; Lu *et al.,* 2012). Finally, the over-use of such compounds can inadvertently select for insect pests with traits that provide resistant to the insecticide used (Feng & Isman, 1995).

1.5 Bio-control methods for aphids and the parasitoid Aphidius ervi

Aphids are predated on by various other, much larger insects such as spiders and the larvae of hoverflies, lacewings, midges, species of Coccinellidae (lady bird beetles). Another common natural enemy of aphids are Hymenopterous parasitoid wasps (Godfray, 1994). Aphid parasitoids have an interesting mode of action, where host organisms are utilized as incubators for the developing parasitoid larvae, devouring the host from the inside and resulting in its death (figure 2). The abundance of parasitoids (and other aphid natural enemies) tend to correlate with the abundance of aphids. During the summer months, the increasing abundance and activity of natural enemies is thought to be the main driver causing declines in aphid numbers (Karley *et al.,* 2003). As in aphids, parasitoid larvae will respond to environmental cues of shortening day length and lower temperatures to enter diapause, overwintering inside the mummified host aphid (Christiansen-Weniger & Hardie, 1997).

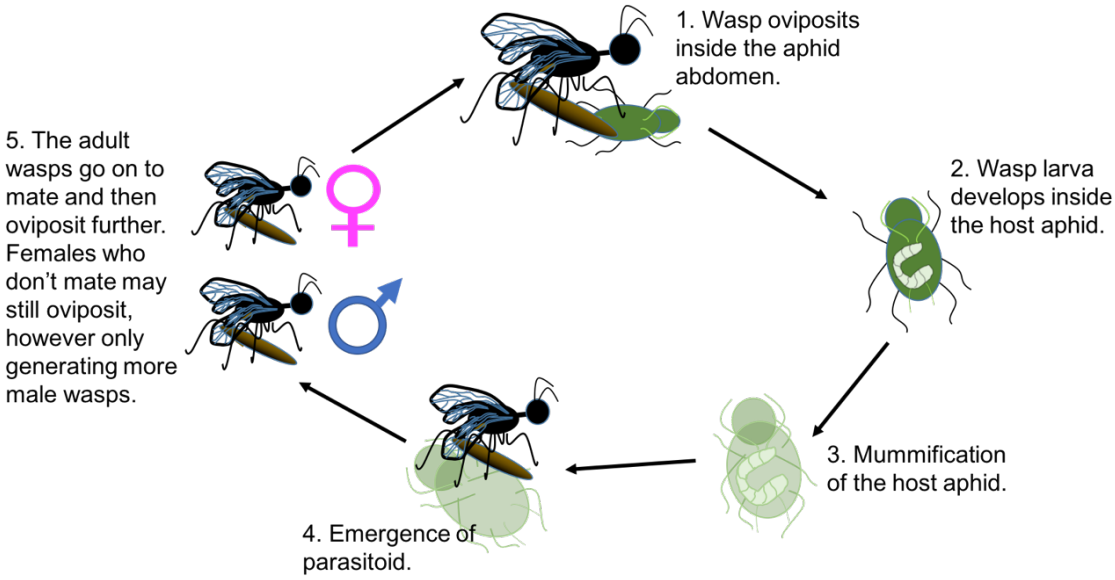


Figure 2. Life cycle of a parasitoid wasp in an aphid. Oviposition will only occur after initial probing and determination of aphid host suitability by the wasp. From oviposition to mummification takes roughly a week, followed by another 4-6 days before emergence of the parasitoid. Rate of development and emergence may differ on environmental factors and wasp species itself.

Hymenopterous parasitoid wasps that target aphids belong to the Braconidae family of parasitoids. With regards to aphids, some of the more common parasitic wasps belong to *Aphidius* (e.g. *Aphidius ervi* (Haliday)) and *Praon* species (e.g. *Praon volucre* (Haliday)) (Jansen, 2005). *Aphidius* wasp species, especially *A. ervi,* are able to attack and oviposit aphids within a matter of seconds, compared to species of *Praon* which take much longer, although they tend to be more efficient, resulting in more successful ovipositions (Sidney *et al.,* 2010). Parasitoid oviposition is also facilitated by the co-injection of wasp venom, containing useful enzymes and compounds that aid in larval establishment within the host (Digilio *et al.,*2000; Colinet *et al.,* 2014) (these are further discussed in Chapter 5). The wasp egg is covered in a sticky chorion outer layer that helps it adhere to tissues within the aphid. In *A. ervi*, this layer breaks after approximately 24 hours to release the serosal morula, consisting of the developing wasp and immature teratocyte cells (Martinez *et al.,* 2016). Teratocyte cells eventually play an important role in hijacking the host metabolism to feed the *A. ervi* larvae (Falabella *et al.,* 2000). Full wasp development within the host takes approximately 10 days (at experimental conditions of a constant 20 °C), with

developmental time also affected by environmental factors or host variations (Sequeira *et al.,* 1992; Zamani *et al.,* 2007).

Natural enemies are efficient at reducing pest numbers (Hickman & Wratten, 1996; Dixon *et al.,* 1997; Michels *et al.,* 2001), thus making them a useful commercial product for aphid and crop pest control. As previously stated however, the effectiveness of these can be reduced through the over-use of chemical pesticides. On the other hand, an over-reliance on biocontrol will have also have drawbacks, as it can be slower to have measurable effects on pest control compared to insecticides, and repeated use selects for individuals resistant to the method of biocontrol as also observed with chemical control methods (Käch *et al.,* 2017; Mills, 2017).

1.6 Role of the aphid obligate symbiont

Aphids are a classic example of mutualism between a eukaryotic host organism and bacterial symbionts. Nearly all aphid species harbour the obligate bacterial endosymbiont *Buchnera aphidicola*, which is necessary for aphid survival (Munsen *et al.,* 1991; Shigenobu *et al.,* 2000). The phloem sap diet of the aphid can often provide insufficient levels of essential amino acids for insect growth and development, and these nutrients cannot be synthesised by the aphid itself. Fortunately, this deficit can be supplemented through essential amino acid biosynthesis by *Buchnera* (Shigenobu *et al.,* 2000). *Buchnera* bacteria are localised to enlarged specialised host cells adjacent to the aphid gut and haemolymph known as bacteriocytes (Braendle *et al.,* 2003). Here, they are provided with non-essential amino acids taken from the aphid diet or are biosynthesised in the bacteriocyte (glutamate, glutamine, proline, serine and tyrosine) (Price *et al.,* 2014) to be used as precursors for generating essential amino acids and vitamins (Shigenobu *et al.,* 2000). Studying the presence and absence of genes linked to synthesis of essential amino acids demonstrates the mutualism between *Buchnera* and its aphid host, where *Buchnera* appears the sole provider of tryprophan for example (Douglas & Prosser, 1992). *Buchnera* also lacks a

complete complement of genes for leucine, isoleucine, valine and methionine biosynthesis (Shigenobu *et al.,* 2000), but this is ameliorated through enhanced localised expression of aphid-encoded enzymes within the bacteriocytes (Shigenobu *et al.,* 2000; Hansen & Moran, 2011).

1.7 Resistance to parasitism and other useful traits of symbiosis

Aphid symbiosis is not restricted to *Buchnera*. *B. aphidicola* is referred to as a primary symbiont, due to its obligate nature with regards to host/symbiont survival. Aphids may also harbour secondary or facultative symbionts, gaining sustenance from the aphid host and *Buchnera* symbiont, but providing a benefit that might not be linked to host nutrition. Even though carrying a secondary symbiont can incur fitness costs and deteriorate life history traits (Heyworth & Ferrari, 2005), they also confer increased tolerance to numerous external stimuli (table 1).

Table 1. Summary of some host benefits associated with harbouring secondary symbionts. '*' Even though *Buchnera* is not a secondary symbiont, it has also been linked to thermal tolerance.

Phenotype	***associated symbiont***	***reference***
Aphid thermal tolerance	*Serratia symbiotica*	Montllor *et al.*, 2002
	**B. aphidicola*	Dunbar *et al*., 2007
Parasitoid resistance	*Hamiltonella defensa*	Oliver *et al*., 2003, Oliver *et al.,* 2005 & Ferrari *et al.,* 2004
	Regiella insecticola	Vorburger *et al*., 2010
	X-type	Heyworth & Ferrari, 2005
Reduced lady beetle survival	*H. defensa/S. symbiotica*	Costopoulos *et al*., 2014
Resistance to fungal pathogen	*R. insecticola*	Scarborough *et al*., 2005
	X-type	Heyworth & Ferrari, 2005
Colour polymorphism	*Ca. Rickettsiella viridis*	Tsuchida *et al.,* 2010

Some of these symbionts play a crucial role in interactions between aphid predators, especially those routinely used in biocontrol methods, one of these being *A. ervi* and the protection provided by the endosymbiont *Hamiltonella defensa* (Oliver *et al*., 2003). The level of protection is strongly suggested to be correlated to the presence of a bacteriophage known as *Acyrthosiphon pisum* secondary endosymbiont bacteriophage (APSE) (van der Wilk *et al.,* 1999), of which there are multiple strains (Degnan & Moran, 2008; Brandt *et al.,* 2017). APSE types vary in the types of eukaryotic toxin homologues they encode, with different toxins believed to underpin variation in the level of symbiont-encoded aphid resistance to *A. ervi.* For example, lines encoding homologues to YD-repeat toxins provide a strong killing action against the wasp larvae, while those carrying cytolethal-distending toxin genes only provide moderate protection (Moran *et al.,* 2005; Brandt *et al.,* 2017). Another bacterial-toxin encoded method of resistance is also potentially conferred by *R. insecticola,* although the underlying mechanism

is less well understood (Hansen *et al.,* 2012). Symbionts may also indirectly protect host aphids from parasitism, for example when symbiont presence reduces levels of herbivore induced plant volatiles that act to attract parasitic wasps for protection (Frago *et al.,* 2017). Other examples include colour polymorphisms as a result of symbionts (Tsuchida *et al.,* 2010) that may affect parasitoid choice, with green aphids preferentially attacked over other colour morphs (Libbrecht *et al.,* 2007), or symbionts that result in behavioural changes that modify aphid responses during challenge by *A. ervi* (e.g. ameliorates loss of feeding opportunities that occur during challenge) (Dion *et al.,* 2011).

1.8 The role of monitoring in IPM strategies

Monitoring of individuals within a population is another important facet of IPM strategies, whether though visual monitoring (Coli *et al.,* 1985) or more commonly through genetic screening for diagnostic markers. For example, monitoring is used in anti-malarial programs, where *Anopheles* mosquitoes are assessed for the presence of knockdown resistance point mutations that confer insecticide resistance (Henry-Halldin *et al.,* 2012; Mavridis *et al.,* 2018). Where biocontrol methods may occasionally have little effect or even fail (Tomasetto *et al.,* 2014; Mills, 2017), monitoring will also provide useful information on IPM strategies efficacy.

Diagnostic screening is also used in IPM strategies for crop pests such as the Peach-potato aphid *Myzus persicae* (Sulzer), where multiple routes to insecticide resistance have occurred (Srigiriraju *et al.,* 2010; Tang *et al.,* 2017). In the UK, major insect sampling is carried out through the Rothamsted Research suction trap network, which provides weekly data for insect surveying and is a useful tool for forecasting pest frequencies (mainly aphids and moths) and prevent unnecessary preventative use of insecticides (Harrington & Woiwod, 2007).

Monitoring of insect pests may also incorporate the use of molecular markers that are diagnostic of individuals that express specific phenotypes, and aid the

understanding the population structure at given locations (Franklin *et al.,* 2010; Weng *et al.,* 2010). Discovery of these genetic markers has been further enhanced through recent improvements in whole genome sequencing and the increased information coming from Arthropod genomic and transcriptomic projects (i5K Consortium, 2013; Legeai *et al.,* 2010). For example, studies into the Pea aphid *Acyrthospiphon pisum* (Harris) have highlighted differentiation in genomic loci correlated with aphid survival on either alfalfa, pea or clover, where loci are adjacent to salivary protein encoding genes or those encoding olfactory receptors (Jaquiéry *et al.,* 2012). Other examples include the transcriptomic analyses of cryptic species within the invasive whitefly *Bemisia tabaci* (Gennadius) and the identification of positively selected genes for metabolism (related to host preference) and insecticide resistance between two spatially-separated populations (Wang *et al.,* 2011); and the generation of genomes for *Anopheles* malaria mosquitoes to elucidate mechanisms behind insecticide resistance as well as provide molecular markers for their identification within the population (Donelly *et al.,* 2016; Clarkson *et al.,* 2018). This ultimately improves the efficiency of IPM strategies, preventing the use of redundant methods of control or helps determine where specific strategies will provide the most benefit.

1.9 The Potato aphid Macrosiphum euphorbiae

Macrosiphum euphorbiae (Thomas) is an economically important pest of potato (as well as other plant crops such as lettuce, cabbage, tomato, raspberries and strawberries) and is found throughout North America and most of Europe. Potato aphid populations confined to North America undergo a sexual phase and overwinter as eggs (MacGillivray & Anderson, 1964), with rose plant species being utilized as primary hosts. However, previous evidence suggests the primary mode of reproduction in European populations of *M. euphorbiae* is strict asexuality, with possibly much rarer occurrences of sexual reproduction (Raboudi *et al.,* 2012). The Potato aphid generates both alate and apterous forms, and similar to the pea aphid has green and pink colour morphs (figure 3). From nymph birth to adulthood takes approximately 9-11 days, going through four nymph instars and

moults beforehand. The apterous forms have a distinctly darker stripe running along their back, similar only to other *Macrosiphum* species. *M. euphorbiae* aphids are predated on by many of the natural enemies commonly associated with aphids, including various parasitoid species. In North America, *Aphidius nigripes* (Ashmead) is the most common aphid parasitoid, whereas *A. ervi* is dominant in Europe, with a recent study showing 54% of all parasitized aphids found across 11 potato fields in Belgium belong to *A. ervi* (Jansen, 2005), although their specific effectiveness against *M. euphorbiae* in this setting is questioned.

A)

B)

Figure 3. Potato leaves infested with either green (A) or pink (B) colour morphs of *M. euphorbiae*. Genes underlying colour polymorphism are likely to be carotenoid biosynthesising genes horizontally transferred from fungi, and is described further in the Pea aphid by Moran & Jarvik (2010).

While *M. persicae* is considered the major contributor to yield loss and crop damage in potato compared to other aphid species, *M. euphorbiae* still has the potential to reduce yield on certain potato varieties. This is discussed in Veen (1985), where infestation on potato plants resulted in carbohydrate accumulation and leaf rolling, in turn reducing yield in the tuber. However, aphid infestation rarely reaches sufficiently high levels in UK potato crops to cause significant direct damage (Parker, 2005). *M. euphorbiae* typically causes crop damage as a vector of multiple plant viruses, such as potato virus Y and potato leaf roll virus, while also transmitting pathogenic viruses that affect beans, sugar beets, sugarcane and lettuce for example (Blackman & Eastop, 2000). Potential spread is further enhanced through the polyphagous nature of the Potato aphid, as it can survive on over 200 plant species, covering more than 20 taxonomic families (Blackman & Eastop, 2000; Srinivasan & Alvarez, 2010).

1.10 Current IPM strategies regarding M. euphorbiae

Due to the polyphagous ability of *M. euphorbiae*, control of this species is incorporated into IPM programs for a range of commercial crops, with biocontrol playing and important role. This includes the use of UV radiation blocking mesh to reduce abundance of *M. euphorbiae* in lettuce (Legarrea *et al.,* 2012); the use of predatory heteroptera on tomato crops (Alvarado *et al.,* 1997); and various parasitoid species in strawberries (Sampson *et al.,* 2011). With regards to potato, chemical methods of control are still routinely used to curb aphid numbers, even though there is some suggestion of insecticide resistance in some clonal lines of *M. euphorbiae,* specifically to carbamates, organophosphates and pyrethroids (Foster *et al.,* 2002). The genetic basis of resistance is studied extensively in *M. persicae* and relates to amino acid substitutions in acetylcholinesterase (AChE), which is involved in neurotransmission (Bass *et al.,* 2014), with these point mutations reducing sensitivity to insecticidal compounds. The risk of selecting for insecticide resistance in populations of the Potato aphid further highlights the requirement for alternative modes of control, such as those previously

mentioned. The controlled release of parasitoids may be one facet by which IPM for *M. euphorbiae* can be improved.

Not only are *M. euphorbiae* in the UK suggested to persist primarily via obligate asexuality, current information suggests the number of distinct genotypes that exist within the UK is small. Of these few genotypes, a specific genotype of Potato aphid shows potentially higher esterase activity (normally linked to the detoxification and removal of insecticide based compounds) (Foster *et al.,* 2002; Clarke *et al.,* 2018), and presents high levels of resistance against the common parasitoid wasp, *A. ervi* (Clarke *et al.,* 2017). Normally, attack by *A. ervi* and successful oviposition of a wasp larvae results in death and mummification of the host aphid, where the developing larvae devours the host from the inside (see figure 2). However, successful mummification often fails to materialise in this genotype of *M. euphorbiae*. Parasitoid resistance in aphids is frequently associated with the presence of facultative endosymbionts (see table 1), but resistance in this instance appears innate and aphid-encoded. Some studies have observed some effect of aphid clonal genotype on *A. ervi* resistance (Li *et al.,* 2002; Martinez *et al.,* 2014), but the genetic reason behind it is not known.

1.11 Innate resistance to parasitism in other insects

Innate resistance to parasitoids remains difficult to study, even in model *Drosophila* species where it appears to be a complex trait mediated by multiple genes and pathways (Kim-Jo *et al.* 2019). Selection for resistance can even arise through multiple routes, as highlighted by Gerritsma *et al.* (2019) where laboratory lines of *Drosophila melanogaster* showing decreased susceptibility to the parasitoid *Asobara tabida* harbour different signatures of selection for resistance compared to natural field populations of *D. melanogaster* that have heightened immunity to the parasitoid (also see Jalvingh *et al.,* 2014). Work by Wertheim *et al.* (2005) suggests a role for genes previously unrelated to immunity, as well as demonstrates how the defence response differs from that observed in anti-microbial immunity. This again suggests how poorly

characterized parasitoid resistance is and how complicated it can be. Biochemical assays have, however, highlighted some mechanisms of interest, specifically those linked to the cellular immune response and encapsulation of foreign bodies (Kacsoh & Schlenke, 2012; Schmitz *et al.,* 2012). In *Drosophila*, one of the main components of the encapsulation response are genes that encode pro-phenoloxidase (PPO) (Dudzic *et al.,* 2015). PPO is cleaved by serine proteases (serpins) to phenoloxidase (PO) which in turn generates melanin from phenolic compounds. Melanin is deposited on and around the body of the invader to aid in killing by host immune haemocytes (Schmitz *et al.,* 2012; Dudzic *et al.,* 2015). Phenoloxidase activity is maintained by this cellular immune response. In the Pea aphid, five morphological types of haemocytes have been identified, of which two (granulocytes and spherulocytes) have observed invader-killing behaviour (Schmitz *et al.,* 2012). Granulocytes are involved in the breakdown of invading organisms while spherulocytes are responsible for coagulation of other haemocytes and pathogens.

What remains intriguing about innate resistance in *M. euphorbiae,* however, is the apparent lack of immune system responses associated with some aphid species. Pea aphids demonstrate reduced immune-function pathways, lacking genes commonly associated with IMD/JNK signalling and the induction of antimicrobial peptides as well as peptidoglycan recognition receptors for identifying foreign bodies (Gerardo *et al.,* 2010). It is also suggested that while still somewhat active, the PO response and melanisation cascade is attenuated to allow the presence of useful endosymbiotic bacteria (Douglas *et al.,* 2010), but in turn impairs the mounting of a successful immune response to bacterial infections or parasitic invaders, such as developing parasitoid larvae.

1.12 Aims and objectives

The main goals of this study are to further investigate specific aspects of the biology of *M. euphorbiae* to understand the traits that underlie its status as a crop pest. It has been established that a specific genotype of the Potato aphid presents high levels of resistance to parasitism by the common parasitoid *A. ervi*, posing a threat to current IPM strategies. Therefore, this thesis will focus on *A. ervi* resistance in potato aphid populations in order to:

1. Describe the genotypic structure of potato aphid populations over areas of Scotland and Merseyside and characterise the frequency of the parasitoid resistant genotype (Chapter 3);
2. Investigate variation in aphid fitness in specific life-history traits of common potato aphid genotypes to further understand the observed genotypic composition of aphid populations and any cost-benefit trade-offs associated with parasitoid resistance (Chapter 3);
3. Assemble and annotate a high-quality near-chromosome scale genome of parasitoid resistant genotype of *M. euphorbiae* and the associated *B. aphidicola* endosymbiont using long-read sequencing and current long-range scaffolding techniques (Chapter 4); and
4. Predict putative causes of parasitoid resistance using inter-species and inter-genotype comparative genomics to identify candidate genes underpinning resistance alongside analysis of genotype-specific patterns in the aphid metabolome (Chapter 5).

The outcome of this work aims to inform IPM for *M. euphorbiae*, addressing if parasitoid resistance is on the rise in the UK. Genome sequencing of multiple genotypes of *M. euphorbiae* will also provide a resource of genomic data, and may be able to provide targets for future breeding programmes for *A. ervi* that are able to overcome innate aphid resistance.

2. General methods

While each section of this work required its own specific protocols and workflows, there was also overlap in some of these methods, mainly regarding the maintenance of insect cultures and molecular work. This brief chapter outlines these methods for future reference by experimental chapters 3, 4 and 5.

2.1 Insect culture and maintenance

Aphids were kept in enclosed ventilated cups (5 inches tall, 3.5 inch diameter) under constant conditions 16h:8h L:D, 20 °C ± 1 and 70% RH on leaf cuttings of potato (cv. Desirée) with the cut petiole placed in water. Potato plants were grown in on-site greenhouses. Mummies of the parasitoid *Aphidius ervi* were purchased from Fargro (West Sussex, UK) and kept at constant conditions of 16h:8h L:D, 20 °C ± 1 and 70% RH. Emerging wasps were fed using a cotton ball soaked in 50% honey. *A. ervi* cultures were maintained separately from aphid cultures to prevent parasitism in stock clonal lines.

2.2 Genomic DNA extraction and preparation of short read Illumina libraries

Five to ten adult aphids were flash frozen in liquid nitrogen and homogenised with a plastic pestle in a 1.5 mL tube. The DNA extraction was performed using Zymo gDNA miniprep Kit (California, USA), following the manufacturer's instructions. An additional RNase treatment was added, in which samples were incubated with RNase (Sigma-Aldrich, Missouri, USA) at 37 °C for 30 minutes, followed by clean-up with a standard AMPure beads protocol (Beckman Coulter, California, USA) (1:1 ratio of sample to beads). DNA integrity and quantity was assessed using Nanodrop and Qubit (Thermofisher, Massachusetts, USA) respectively and sized using a 0.5% agarose gel run overnight at 26 V; 500 mA. DNA size was measured against the NEB 1kb extend ladder, with a top band of 40 kb.

2.3 M. euphorbiae challenge with the parasitoid Aphidius ervi

Approximately 20 adult aphids were left to deposit nymphs for four days on an excised potato leaf in an enclosed ventilated cup under constant conditions 16h:8h L:D, 20 °C ± 1 and 70% RH on leaf cuttings of potato (cv. Desirée). From these, 20 nymphs aged approximately four days old were transferred to a new excised leaf under the same conditions. 4-day-old nymphs were placed carefully on a cut leaf embedded in agarose in a petri dish and left to settle for a minimum of 30 minutes. *A. ervi* cultures were anaesthetized briefly using CO_2 and a single female between 2-5 days old (presumed mated) was placed into the petri dish with the nymphs. After a single attack (when the wasp ovipositor was observed to insert into the aphid abdomen), each nymph was moved to another potato leaf with its stem embedded in agarose in a petri dish and returned to constant conditions previously described. The process was repeated until 15 nymphs had been attacked once for each genotype assessed. If wasps were unresponsive for 5 minutes, they were replaced with another wasp.

2.4 Data sources used in analysis

All data used for work in this thesis and web links are detailed in the online supplementary table found at http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/data_sources_for_PhD.csv. Briefly, all aphid genomes were taken from AphidBase (Legeai *et al.,* 2010) apart from *Acyrthosiphon pisum* and *Rhoplasmosiphum maidis* genomes with integrated Hi-C linkage information, which were taken from NCBI. Protein data sets were taken from AphidBase, with proteins for *R. maidis* taken from NCBI. *Drosophila melanogaster* and Swissprot protein databses were downloaded from UniProt.

3. Population structure insights into *M. euphorbiae*

Abstract

Genetic distribution of aphids can be governed by phenotypes belonging to specific clonal genotypes, especially when such a trait confers a selective advantage. However, these traits can also entail a cost, leading to cost-benefit trade-offs and reducing such the selective advantage in some conditions. The Potato aphid (*Macrosiphum euphorbiae*) is suggested to persist as a handful of asexual genotypes throughout the UK. A specific genotype of this aphid species exhibits strong resistance to the hymenopterous parasitoid wasp *Aphidius ervi*. However, little is known about the abundance or other phenotypic traits of this aphid genotype. From 132 samples collected across three years and two regions (Dundee in Scotland and Merseyside in England), 4.5% belonged to the parasitoid resistant genotype 1, with the majority belonging to previously characterised as genotypes 2 (34.1%) and 3 (25.8%). Strict clonality was further supported through STRUCTURE analysis and the separation of clones based on genotype rather than sampling location. Interestingly, while genotype 1 has low abundance in the population, there was no obvious adverse effect of parasitism resistance on other aspects of aphid fitness. The rate of aphid development and fecundity following aphid challenge/parasitism by *A. ervi* was unaffected in genotype 1 aphids and genotype 2 aphids that survived *A. ervi* challenge, but fitness was decreased in surviving genotype 3 aphids. In fact, adults of genotype 1 demonstrated slight increased survival in response to a cold shock compared with adults of genotypes 2 and 3. This study is the first to report the prevalence of parasitoid resistant genotypes within UK *M. euphorbiae* populations, and demonstrates a positive association between two traits that increase survival under biotic (parasitism) and abiotic (temperature) stresses.

3.1 Introduction

3.1.1 Phenotypic and genotypic variance within aphid species

Insect pest populations can vary phenotypically and genotypically. Therefore, being able to monitor the dynamics of these pests can be useful tool for targeted and appropriate insect pest management (IPM) responses, with aphids being a common example of this (Bournovile *et al.,*2000; Miller *et al.,* 2003). Aphid species typically comprise of distinct asexual clonal genotypes that exist throughout the summer months on secondary host plants. Most aphid species generate sexual forms in the autumn, which mate and produce eggs that overwinter on the primary host plant (holocyclic lifestyle) (Simon *et al.,* 2002), although some aphid species persist asexually through the winter (anholocyclic lifestyle) (Dixon, 1973; Simon *et al.,* 1999). Most aphid species overwinter as eggs, while around 30% are believed to consist of a mix of holocyclic and anholocyclic variants with a small proportion (approximately 3%) existing strictly though parthenogenesis (Simon *et al.,* 2002). Studying the genetic structure of aphid populations can reveal factors that determine aphid spread and migration. Aphid genotypes can often be spatially separated, based on geographical range (Guillemaud *et al.,* 2003; Raboudi *et al.,* 2005; Wang *et al*., 2016). In the case of *Aphis gossypii* Glover, spatial structure is observed even over short distances (< 1 Km) between greenhouses (Fuller *et al.,* 1999). Host plant range can also play a large role in aphid spread and frequency. For example, certain *Acyrthosiphon pisum* Harris (pea aphid) genotypes show greater fitness on alfalfa (*Medicago sativa*) or red clover (*Trifolium pretense*) in the wild (Via, 1991; Ferrari *et al.,* 2006), and host plant availability could influence genotype spread. In aphid species that maintain holocyclic methods of reproduction, host specialisation may drive long term speciation by reducing geneflow between these populations (Ferrari *et al.,* 2006).

Monitoring the spread of different aphid genotypes is important for crop growers, especially where specific genotypes can differentially affect plant development (Turley & Johnson, 2015), or show differential efficiency in virus transmission (Yang *et al.,* 2008) or insecticide susceptibility (Devonshire *et al.,* 1998). Insecticide resistance in aphids arguably poses a greater threat to crop growers, with resistance to common insecticides detected in several economically-important aphid species. Genotypes of *M. persicae* have acquired genetic mutations to insecticide targets that result in increased aphid survival in the presence of multiple insecticides, mainly through modification of signalling pathways at synapses within the nervous system (Devonshire *et al*., 1998; Bass *et al.,* 2014).

3.1.2 Genotypic variance and impact on general aphid fitness

Genotypic variation can underlie within-species differences in several aphid fitness traits. The ability to tolerate low temperatures is especially important for anholocyclic aphids to survive overwinter (Dixon, 1973). Interestingly, mode of reproduction is not necessarily indicative of adaptation to cold, as studies suggest that summer morphs of sexually-reproducing lineages could be on par (or better) at surviving cold temperatures than asexual lineages (Bale *et al.,* 1994; Vorburger, 2004). Other factors involved in surviving extreme temperatures include: age of insect, as younger nymphs can show greater cold-hardiness (Clough *et al.,* 1990; Powell & Bale, 2004); behavioural response, such as increased contact and attachment to a host plant (Butts *et al.,* 1997; Alford *et al*., 2014); and symbiont infection status, as the facultative endosymbiont *Serratia symbiotica* has been associated with increased fecundity of pea aphids under elevated temperatures (Montllor *et al.,* 2002). Understanding genotypic differences in aphid survival in response to cold could provide useful information in forecasting the dominant insect pests for the coming year, aiding pest control programme (Adams, 1962; Harvey & Martin, 1988).

Other fitness traits that vary between aphid genotypes include susceptibility to natural enemies. Aphid clones that are resistant to pathogens and/or parasitoids have been detected in the Pea aphid (Ferrari *et al.,* 2001; Martinez *et al.,* 2014) and the Potato aphid *M. euphorbiae* (Clarke *et al.,* 2017). These effects can interact with host preference to affect genotype prevalence: for example, clonal resistance to the generalist parasitoid *Aphidius ervi* in the pea aphid resulted in a higher proportion of parasitized aphids on clover compared with alfalfa (Hufbauer & Via, 1999; Li *et al.,* 2002). As presence of natural enemies can be a major factor regulating aphid abundance (Karley *et al.,* 2003), and aphid genotypic variation in protection against any of these threats could help shape the genotypic composition of aphid populations.

3.1.3 Advantageous survival traits may exist in a cost-benefit system

While insecticide resistance in *M. persicae* has clear and obvious benefits, it may also incur fitness costs (Foster *et al.,* 2007; Foster *et al.,* 2010), as has been noted in other insect species as well (Bourget *et al.,* 2007; Cao *et al.,* 2014). Common trade-offs for resistance to pathogens and chemical insecticides are changes in aphid developmental time and fecundity. Resistance to mortality factors can also be conferred by associations with facultative endosymbionts. Symbiotic relationships within pea aphids are well studied and provide good examples of the costs and benefits of symbiont infection: facultative symbionts can confer increased resistance to parasitoids, entomopathogenic fungi or high temperatures, but the associated costs can include reduced fecundity, shorter lifespan and attenuated defensive behaviours in response to natural enemies (Gwynn *et al.,* 2005; Russell & Moran 2005; Polin *et al.,* 2014; Leclair *et al.,* 2016; McLean *et al.,* 2018).

However, this cost-benefit system is further confounded on some occasions, especially when aphid fitness is increased in response to attack or infection. Altincick *et al.* (2008) describe how wounding and *E. coli* infection in Pea aphids can result in an increase in fecundity, correlating with an increase in wound

healing and lysozyme activity. Regarding parasitoid attack, *Aphis fabae* (Black bean aphid) harbouring the protective *H. defensa* demonstrate increased longevity and fecundity after surviving parasitism by *Lysiphlebus fabarum* compared to unparasitized controls and especially compared to parasitized susceptible clonal lines, resulting in strongly reduced longevity and fecundity (Vorburger *et al.,* 2013). While parasitism may have effects in the attacked host (Vorburger *et al.,* 2013; Kaiser & Heimpel, 2016), increased fecundity and greater overall fitness can be passed-down to offspring of parasitized aphids (Kaiser & Heimpel, 2016). This compensation in response to parasitoid attack or any other exogenous factor could therefore play a role in defining the genetic architecture of a population.

3.1.4 Chapter aims

Although *M. euphorbiae* populations are known to harbour genotypes resistant to the common generalist parasitoid *A. ervi* (Clarke et al., 2017), its frequency in UK populations has not been determined previously. It is important to assess the prevalence of the parasitism-resistant genotype as its presence could impair the success of *A. ervi* in natural populations and hamper the use of *A. ervi* in aphid biocontrol. Further, it is important to quantify any fitness consequences of parasitism resistance that might influence aphid success and persistence.
This study was devised to address the following objectives:

1. Sample native *M. euphorbiae* populations to assess the frequency of genotypes in three consecutive summers and in two localities: in Merseyside and Dundee. This will test the hypothesis that the trait of *A. ervi* resistance affects aphid genotype 1 frequency compared with other susceptible genotypes.
2. Establish clonal lineages of different aphid genotypes and experimentally quantify aphid fitness in terms of resistance against *A. ervi*, and aphid fitness (development, fecundity and survival) after parasitoid oviposition. Aphid development is expected to be perturbed by the effect of the wasp

venom (Falabella *et al.,* 2007), so fitness post attack in genotype 1 will be assessed and compared with susceptible genotypes.

3. Assess the survival of aphid genotypes in response to sub-zero temperatures as a potential fitness cost to parasitoid resistance in genotype 1. Survival in response to cold may help explain variation in genotype frequencies in aphid populations.

The information gathered on genotype frequency, fitness traits and cold tolerance will assist in understanding genotype prevalence and distribution in UK populations of *M. euphorbiae*.

3.2 Methods

3.2.1 Sampling strategy for M. euphorbiae

Over three field seasons, potato aphids were collected from private allotments and commercial farms around the Merseyside (figure 1) area as well as the Tayside, Perthshire and Fife areas of Scotland (figure 2) (table 1). Aphid sampling was carried out to determine the frequency of parasitoid resistance within *M. euphorbiae* populations, and to provide material to study parasitoid resistance and other aphid fitness traits. For commercial farms, between three and five individual were collected each separated by at least 50 paces to avoid sampling aphids that had reproduced clonally in the locality. This approach was not possible at allotment sites, and instead individuals were collected from different allotment plots. In 2016 and 2017, collected aphids were kept at constant conditions of 16h:8h L:D, 20 °C $\pm$ 1 and 70% RH to until determined to be healthy and free of parasitoids and fungal pathogens. In 2018, aphids were immediately frozen and stored at -80 °C until analysis.

Figure 1. Locations of Merseyside sites used for aphid sampling.

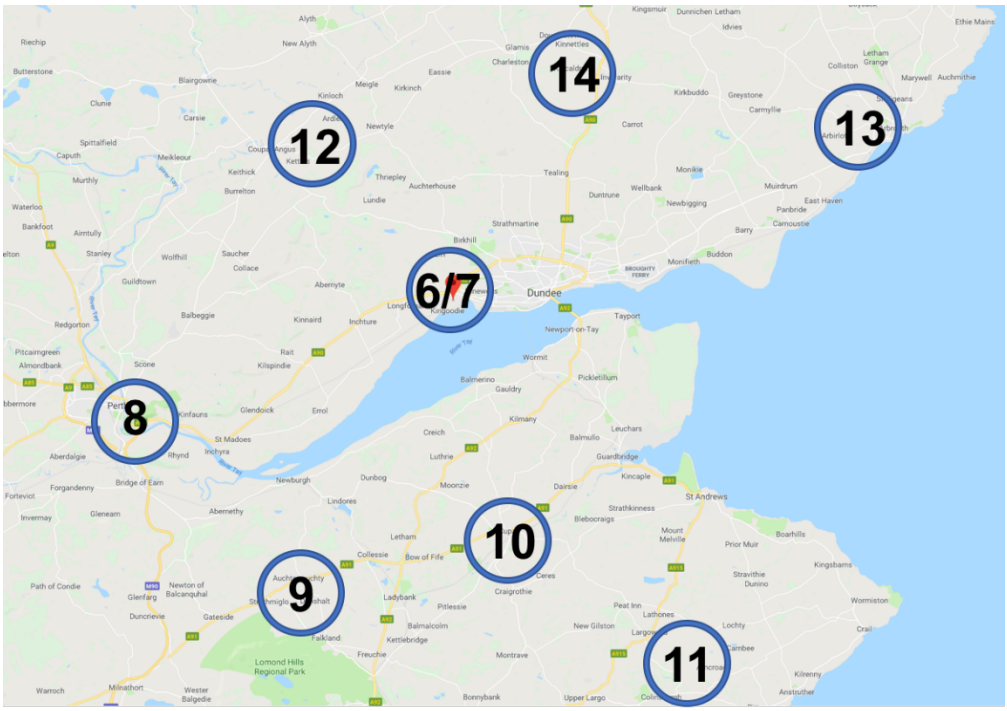


Figure 2. Locations of sites in Tayside, Perth and Fife used for aphid sampling.

Table 1. Characteristics of aphid sampling sites for the potato aphid *Macrosiphum euphorbiae* in the three years of study. See figures 1 and 2.

Area	*Site*	*Type*	*Map key*	*Years visited*
Merseyside	Heswall, Wirral	plot	1	2016, 2017
	Thingwall Allotments, Liverpool	plot	2	2017
	Prenton, Wirral	plot	3	2016, 2017, 2018
	Burton, Cheshire	plot	4	2018
	Claremont Farm	farm	5	2016, 2017
Tayside	Mylnefield, Dundee	plot	6	2016, 2017, 2018
	Balruddery, Dundee	farm	7	2016, 2017, 2018
Perth	Moncrieffe Island Allotments	plot	8	2016, 2018
Fife	Weir Farm	farm	9	2016
	Foodie Farm	farm	10	2016
	Abercrombie Farm	farm	11	2016
Angus	Cardean Field	farm	12	2016
	Brax Farm	farm	13	2016
	Kinreich Farm	farm	14	2016, 2018

3.2.2 Insect culture and maintenance

M. euphorbiae and *A. ervi* were cultured as outlined in section 2.1.

3.2.3 DNA extraction for genotyping and symbiont infection

Between 1-5 aphids of each clonal lineage were flash frozen with liquid nitrogen and ground up in a 1.5 mL tube with a pestle. For samples collected in 2016 and 2018, DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Hilden, Germany). In 2017, the Zymo gDNA miniprep Kit (California, USA) was used. In both cases, DNA was extracted following the manufacturer's protocol. DNA quality and quantity was assessed using Nanodrop (ThermoFisher, Massachusetts, USA).

For surveying symbiont infection status in the clonal lines used in fitness assays, three to four aphids from each line were homogenized and DNA was extracted using the Qiagen DNeasy blood and tissue kit, following manufacturer's instructions. Purified DNA was assessed for symbiont presence by polymerase chain reaction (PCR) and amplification of the 16S rDNA gene and 16-23 S region (specific region absent in *Buchnera*), where 16-23 S positive results indicate secondary symbiont presence. Primers specific for *Hamiltonella defensa, Serratia symbiotica* and *Regiella insecticola* are outlined in sup. Tables 4, 5 and 6. PCR amplicons were visualized using gel electrophoresis and compared against an *Escherichia coli* positive control.

3.2.4 Genotyping collected aphid lines

Up to eight different genotypes of *M. euphorbiae* have been detected amongst UK-collected insects in a previous study (Clarke et al., 2017). Genotyping was performed using restriction fragment length polymorphism (RFLP) analysis of seven microsatellite loci for samples collected in 2016 (*Me1, Me5, Me7, Me9, Me10, Me11* and *Me13*). Only four marker sites (*Me1, Me5, Me9* and *Me10*) were used for samples collected in 2017 and 2018 to streamline analysis and reduce costs, as data from 2016 samples showed these marker sites were sufficient to differentiate the clonal lineages (sup. figure 1). Microsatellite loci were amplified via PCR (sup. tables 1, 2 and 3). Lengths of the resulting PCR products were measured via capillary electrophoresis on a 3730 DNA Analyzer (ThermoFisher, Massachusetts, USA) using the GeneRox 500 internal standard. Microsatellite lengths were assessed using the Genemapper 5 software.

3.2.5 Experimental aphid lines

For the following experiments, a set of six aphid clonal lines were used that had been collected both before and during this project (table 2).

Table 2. Clonal lines of *M. euphorbiae* used in this study.

Clonal Line	***Genotype***	***Collection location***	***Experiment***
MW16/67	1	Fife – 2016	Parasitoid attack, cold exposure
MW17/28	1	Mylnefield – 2017	Cold exposure
MW16/48	2	Perth – 2016	Parasitoid attack, cold exposure
AK15/01	2	(Alford, Lincs - 2015)	Cold exposure
MW16/52	3	Perth - 2016	Parasitoid attack, cold exposure
MW17/04	3	Liverpool – 2017	Cold exposure

3.2.6 Confirming parasitoid susceptibility in 2016 clonal lines

A selection of aphid lines collected in 2016 were assessed for parasitoid susceptibly (table 3). Approximately 20 adult aphids were left to deposit nymphs for 4 days on an excised potato leaf in an enclosed ventilated cup (see section 2.1). From these, 20 nymphs aged approximately 4 days old were transferred to a new excised leaf under the same conditions. A single male and female *Aphidius ervi* wasp between 2-5 days old were transferred into the container with the nymphs for 6 hours, then removed. This was performed in triplicate for each line tested. After one week, the number of aphid mummies was counted. After another week, the number of wasps that emerged from these mummies was also counted.

Table 3. Established clonal lineages assessed for parasitism resistance. '*' denotes parasitoid resistant genotypes. '+' denotes llineage belongs to a pink variety, with the remainder being green. Lines AK13/30 and RB15/10 were established at the James Hutton Institute prior to this study.

Clonal Line	***Genotype***	***Symbiont status***
AK13/30	2	*H. defensa + APSE*
MW16/06	6	*H. defensa + APSE*
MW16/113	7	-
MW16/38	2	*H. defensa + APSE, R. insecticola*
MW16/40	6	-
MW16/48	2	-
MW16/52	3	-
MW16/67*	1	-
MW16/88	3	*H. defensa + APSE*
MW16/95	3	*H. defensa*
MW16/98	p7 $^{+}$	*H. defensa + APSE*
RB15/10*	1	*H. defensa + APSE*

3.2.7 Assessing aphid response to parasitoid attack

Wasp assays were performed as outlined in section 2.3. The control treatment comprised of 15 non-attacked aphids for each genotype tested, otherwise treated in the same way. Aphids were transferred to a new leaf embedded in agarose approximately once a week to maintain healthy plant material in the Petri dish. Aphid survival was monitored over a 30-day period, measured from the day of nymph birth. The dates for development into adulthood, first reproduction, death and mummification were recorded. Average nymph production per day was calculated from when an individual reached adulthood. Life history traits were analysed from individuals that survived parasitism only. Measurements were collected in two separate experiments comparing genotypes 1 and 3, followed by genotypes 1 and 2.

3.2.8 Aphid survival in response to cold temperatures

Two clonal lines were selected of genotypes 1, 2 and 3 (previously assessed using microsatellite analysis) (table 2) that were shown to be free of secondary symbionts using diagnostic PCR, knowing that symbionts can potentially confer thermal tolerance (Montllor *et al*., 2002). Adult aphids maintained at 16h:8h L:D, 20 °C $\pm$ 1 were used to generate cohorts of aphids of known age: 10 (adult) days, 7 days (3rd instar) and 3 days old (2nd instar). Three days prior to the assay, culture temperature was dropped to 10 °C $\pm$ 1 (16h:8h L:D). The following day, aphids were moved to a 5 mL Eppendorf^{TM} tube containing a small cutting of potato leaf stem embedded in agar, with 10 aphids per replicate, and 10 replicates of each genotype/age combination for both control and experimental procedures. The tubes containing the aphids were placed into an incubator in a randomized block design. Four incubators were used for this experiment: two for the cold exposure treatment and two for the control treatment. Controls were maintained at 10 °C (16h:8h L:D). For cold exposure, aphids were subject to -5 °C for 3 hours, followed and preceded by an hour at 3 °C. A light regime of 16h:8h L:D was maintained during the experiment, with cold exposure occurring during the 8 hour period of no light. It was impossible to source four identical incubators. Therefore, two sets of identical incubators were used (Snijders Jumo IMAGO F3000 and Snijders Jumo ILP F200) with one of each set for either cold exposure or control temperatures. Aphid survival was measured 36 hours after exposure to cold or control temperatures.

3.2.9 Statistical analyses

Analyses were performed using R (v3.2.2). Graphical output was generated using R and ggplot2. Chi-squared analysis was used to assess genotype distribution over three years of sampling. GLM analysis of count data with Poisson distribution was used to assess the effect of year, sampling location and site type on observed genotype frequencies. Adegenet (v2.1.1) in R was used to generate PCA plots with

microsatellite loci using default settings and visualized in ggplot2. Population structure based on these microsatellites was generated using the Bayesian clustering algorithm STRUCTURE (v2.3.4) (Pritchard *et al.,* 2000). STRUCTURE analysis demonstrates genetic relationships between genotypes and evidence of any recombination before the switch to parthenogenesis. Parameters include a Burnin period of 10,000 with 50,000 MCMC replicates after Burnin. For *K* of 3 to 24 (with 24 being the theoretical maximum based on unique genotypes described here), 10 replicates were performed, where *K* is the number of distinct genetic groups. STRUCTURE uses an admixture model to account for possible genetic recombination between clones before the switch to parthenogenecity. Admixture can be turned off, but is recommended where possible genetic recombination is unknown (Falush *et al.,* 2003). STRUCTURE output was uploaded to STRUCTURE HARVESTER (web v0.6.94) (Earl & vonHoldt, 2012), an online tool to predict the best fitting value of *K* produced from STRUCTURE using the Evanno delta *K* method (Evanno *et al.,* 2005).

Aphid survival in response to parasitoid attack of the 15 clonal lines was assessed using one-way ANOVA to confirm increased parasitoid resistance in genotype 1 samples. Aphid development variables in response to parasitism were assessed with ANOVA and Tukey's HSD. Survival plots were generated using the survminer package (0.4.2). Statistical analyses (log rank test) were performed in R using survminer. Normality plots of residuals for each ANOVA were performed in R to confirm normal distribution for all data. When comparing the two experiments comprising of genotypes 1 and 2 and genotypes 1 and 3, MANOVA was used to compare genotype 1 factors between the two experiments to decide if analysis should be combined or kept separate.

For cold exposure experiments, a randomised block design was generated using the R package Agricolae v1.2.8 (Mendiburu, 2018). Significant differences in survival counts were assessed using two generalized linear mixed models (GLMM) with binomial error structure, measuring the interaction between age (3, 7 and 10 days old), treatment (cold exposure or control) and clonal line or genotype. Initial

models nesting clonal line within genotype resulted in a rank deficient model. Therefore, the model was simplified by analysing for the factors of clonal line or genotype separately. GLMM was generated using glmer as part of Lme4 (Bates *et al.,*2014).

3.3 Results

3.3.1 Genotype distributions of M. euphorbiae

Over three years, 132 individual aphid collections were genotypically characterised through molecular analysis of four microsatellite positions, with marker sizes and genotype identity matched those demonstrated in a previous study (Clarke *et al*., 2017). Chi-squared analysis showed genotype proportions were not equally represented in the tested populations (χ^2 = 377.45, df = 17, P < 0.001). The majority of sampled aphids belong to green clones, with genotype 2 being the most common (n = 45), followed by genotype 3 (n = 34). Genotypes 6 (n = 11) and genotype 7 (n = 5) were less abundant, as was the parasitoid resistant genotype 1 (n = 6). PCA analysis based on these microsatellite loci demonstrated that genotypes 2 and 3 are the most distantly related, while genotypes 1 and 6 may bear some genetic resemblance (figure 3). 12 green individuals could not be assigned to a previously characterized genotype. Genotypes 4 and 5 detected in previous study (Clarke et al., 2017) were not observed over three years of sampling. Pink clones accounted for 20 sampled individuals. Green clones are more abundant and better characterized, while pink clonal lineages are less well defined. Within pink clones, three clusters were observed (p1, p2 and p7), while the remaining four clones could not be grouped (figure 4). Variants within some genotypes were observed, where a microsatellite size differed at a single allele. (Microsatellite positions for each genotype and its potential variants are described in http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/genotype_Me_calls.csv.

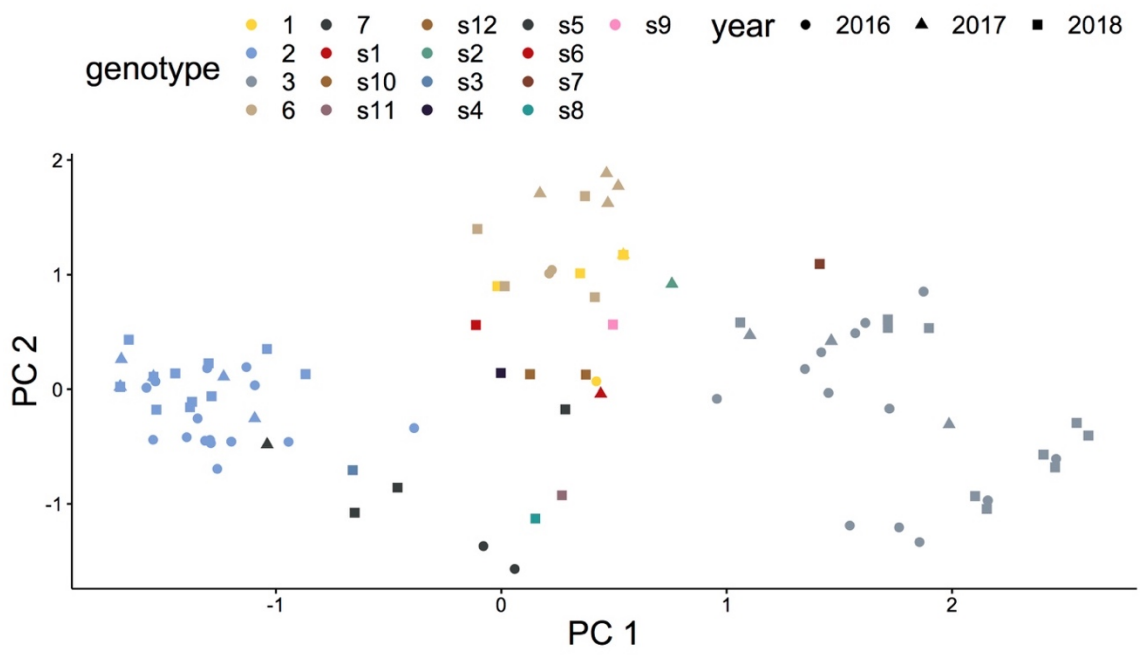


Figure 3. PCA plot of green *M. euphorbiae* genotypes based on 4 microsatellite positions. 's' clones are singletons and unplaced within previously-characterised genotypes. PC1 and PC2 explain 26.3% and 6.9% of the data repesctively.

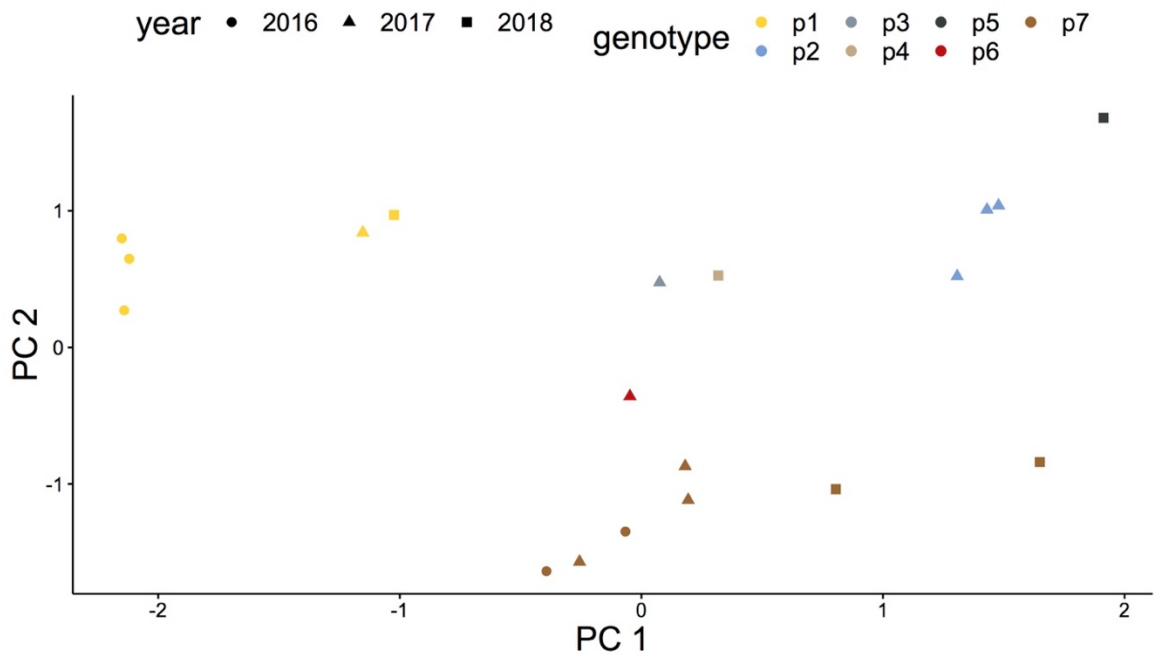


Figure 4. PCA plot of pink *M. euphorbiae* genotypes based on 4 microsatellite positions. p3, p4, p5 and p6 are singletons and cannot be assigned to specific pink genotypes. PC1 and PC2 explain 22.3% and 16.0% of the data respectively.

GLM analysis of count data including geographic location, year and site type (plot or commercial farm) indicated significantly higher abundance of genotypes 2 and 3 compared with other genotypes (table 4). Significant differences in genotype abundance were observed at Tayside (Balruddery and Mylnefield) and Fife sites. While not significant, genotype 6 was most common in Merseyside. Site type was a significant factor in genotype distribution. This effect is driven by the lack of genotype 3 found at commercial farms and only seems present at plot/allotment style settings (table 6).

Table 4. GLM analysis genotype abundance of *M. euphorbiae* at collection sites in Merseyside (England) and Eastern Scotland. Baseline category for GLM analysis was counts for genotype 1 at a commercial site in Angus, 2016.

	Estimate	***Std. Error***	***z value***	***Pr(>|z|)***	
(Intercept)	-2.77614	0.71399	-3.888	0.000101	***
genotype2	2.0149	0.43461	4.636	3.55E-06	***
genotype3	1.7346	0.44281	3.917	8.96E-05	***
genotype6	0.60614	0.50752	1.194	0.232356	
genotype7	-0.18232	0.60553	-0.301	0.763342	
genotypep1	-0.18232	0.60553	-0.301	0.763342	
genotypep2	0.77614	0.72774	1.067	0.286195	
genotypep3	-0.32247	1.09375	-0.295	0.768124	
genotypep4	-0.82557	1.08677	-0.76	0.44746	
genotypep5	-0.82557	1.08677	-0.76	0.44746	
genotypep6	-0.32247	1.09375	-0.295	0.768124	
genotypep7	0.15415	0.55635	0.277	0.781722	
genotypes1	-0.32247	1.09375	-0.295	0.768124	
genotypes10	-0.82557	1.08677	-0.76	0.44746	
genotypes11	-0.82557	1.08677	-0.76	0.44746	
genotypes12	-0.82557	1.08677	-0.76	0.44746	
genotypes2	-0.32247	1.09375	-0.295	0.768124	
genotypes3	-0.82557	1.08677	-0.76	0.44746	
genotypes4	-0.82557	1.08677	-0.76	0.44746	
genotypes5	-0.82557	1.08677	-0.76	0.44746	
genotypes6	-0.82557	1.08677	-0.76	0.44746	
genotypes7	-0.82557	1.08677	-0.76	0.44746	
genotypes8	-0.82557	1.08677	-0.76	0.44746	
genotypes9	-0.82557	1.08677	-0.76	0.44746	
siteplot	0.96161	0.23032	4.175	2.98E-05	***
year2017	-0.08697	0.2757	-0.315	0.75243	
year2018	0.1971	0.23536	0.837	0.402354	
locFife	2.32542	0.65977	3.525	0.000424	***
locMerseyside	0.55687	0.65366	0.852	0.394255	
locPerth	0.93551	0.66057	1.416	0.156711	
locTayside	1.59782	0.61545	2.596	0.009427	**

Table 5. Genotype abundance of *M. euphorbiae* based on geographic location for each year of sampling. N = 44, 33 and 55 for 2016, 2017 and 2018 respectively.

year	*sample location*	*1*	*2*	*3*	*6*	*7*	*p1*	*p2*	*p7*	*unplaced*	*Total*
2016	*Angus*	0	0	0	0	1	1	0	0	0	2
	Tayside	0	7	11	1	1	1	0	1	0	22
	Fife	1	10	0	0	0	0	0	1	0	12
	Merseyside	0	1	0	1	0	0	0	0	0	2
	Perth	0	1	4	0	0	1	0	0	0	6
2017	*Tayside*	1	9	2	0	1	1	3	0	4	21
	Merseyside	0	3	2	4	0	0	0	3	0	12
2018	*Angus*	1	0	0	0	0	0	0	0	0	1
	Balruderry	1	13	6	1	2	0	0	2	10	35
	Merseyside	0	0	1	2	0	0	0	0	2	5
	Perth	2	1	8	2	0	1	0	0	0	14
	Total	**6**	**45**	**34**	**11**	**5**	**5**	**3**	**7**	**16**	**132**

Table 6. Genotype abundance of *M. euphorbiae* on commercial farms and allotment plots for each year of sampling. N = 41 and 91 for farm and plot sites respectively.

year	*site type*	*1*	*2*	*3*	*6*	*7*	*p1*	*p2*	*p7*	*unplaced*	*Total*
2016	*farm*	1	14	0	1	1	1	0	1	0	19
	plot	0	5	15	1	1	2	0	1	0	25
2017	*farm*	0	1	0	3	0	0	1	0	2	7
	plot	1	11	4	1	1	1	2	3	2	26
2018	*farm*	1	8	0	1	1	0	0	1	3	15
	plot	3	6	15	4	1	1	0	1	9	40
	Total	**6**	**45**	**34**	**11**	**5**	**5**	**3**	**7**	**16**	**132**

3.3.2 STRUCTURE analysis supports strict clonality

STRUCTURE was used to assess the presence of shared genetic material between genotypes of *M. euphorbiae*. STRUCTURE and STRUCTURE HARVESTER predicted *K* = 5, as denoted by the highest delta *K* value (sup. figure 2). STRUCTURE output supported the PCA analysis, showing that green genotypes 1, 2, 3 and 6 formed their own clusters, as well as pink genotypes p1, p2 and p7, as can be demonstrated through uniformity in colour within each genotype and distinct colours between genotypes (figure 5). Genetic clustering is not observed at the geographic level, suggesting a lack of recombination and predominance of asexual reproduction in UK populations of *M. euphorbiae*. As *K* is not equal to the number of observed distinct clonal genotypes, it could indicate genetic overlap in genotypes, or be a consequence of microsatellite data not providing enough resolution to classify genotypes correctly.

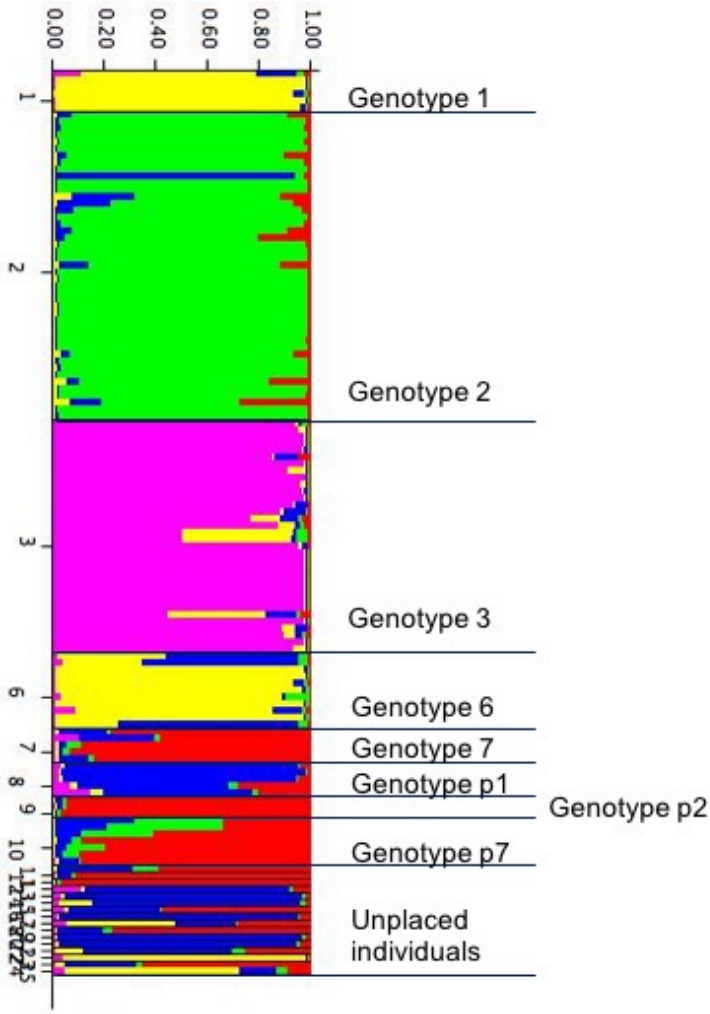


Figure 5. Population structure of *M. euphorbiae* (*K* = 5) generated through STRUCTURE, using 132 individual samples (each represented by a single horizontal bar) and four microsatellite loci per sample. Each colour represents one of five predicted populations, while each bar comprises of coloured segments representing admixture proportions from all populations of *K*.

3.3.3 Preliminary resistance assays confirm high levels of resistance in samples of genotype 1

Genotype 1 aphids showed a high level of resistance to parasitism by *A. ervi* in parasitism assays (figure 6). One-way ANOVA confirmed significantly increased survival of Genotype 1 aphids after parasitoid attack ($F_{5,38} = 41.93$, $P < 0.001$). Tukey's HSD post-hoc analysis showed significantly higher survival of genotype 1 aphids compared with all other genotypes tested ($P < 0.05$). One replicate from the genotype 2 assay MW16/48 was omitted due to observing no aphid death or parasitism, possibly due to the presence of a drowned female *A. ervi* in the bottom of the assay cup, resulting in no attack.

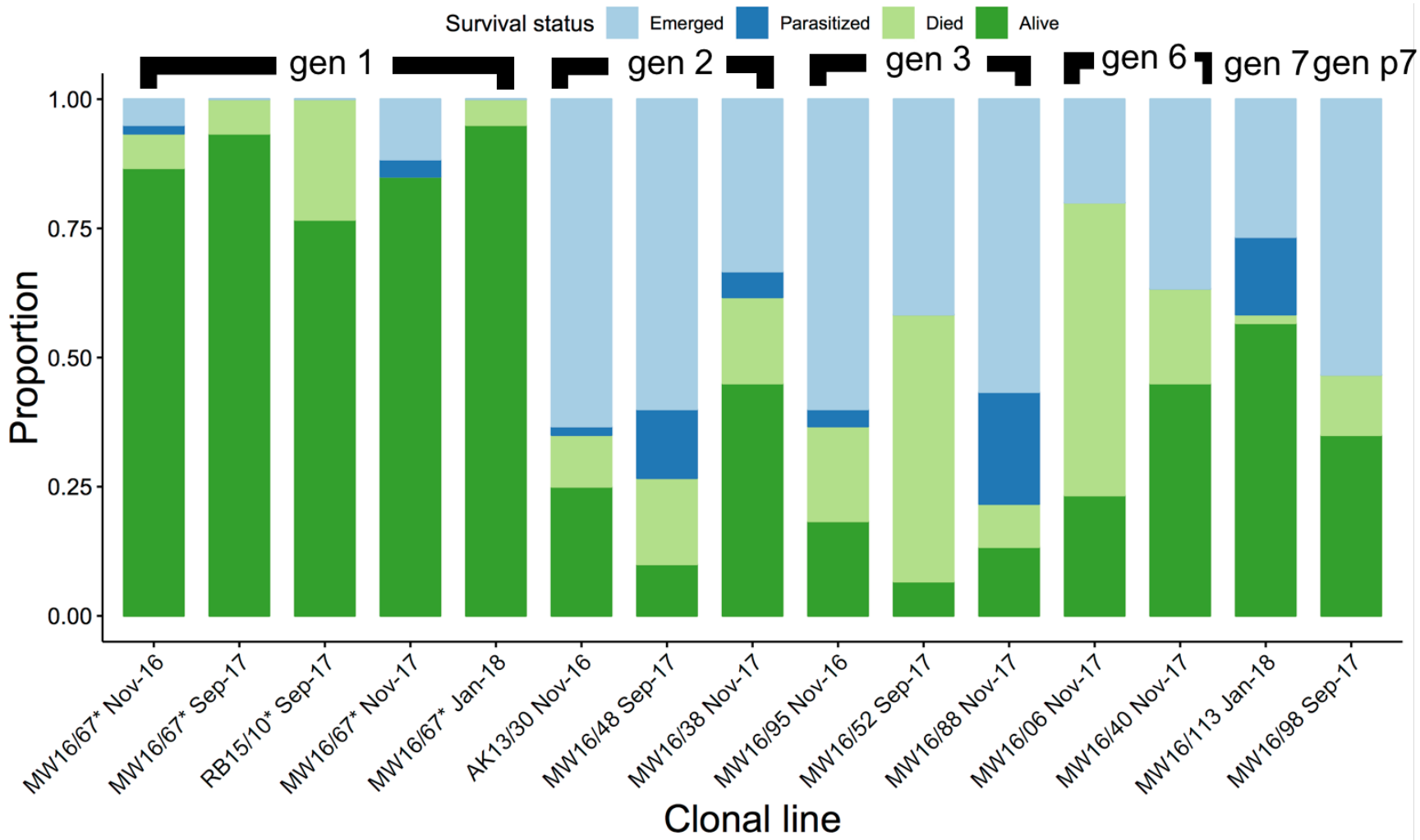


Figure 6. Success of *A. ervi* parasitism (proportion of aphids mummified) amongst aphid genotypes sampled in 2016 (present study) and collected prior to this study. Values are means of three replicates with 20 nymphs per replicate. Date indicates time of testing. '*' lines belong to parasitoid resistant genotype 1.

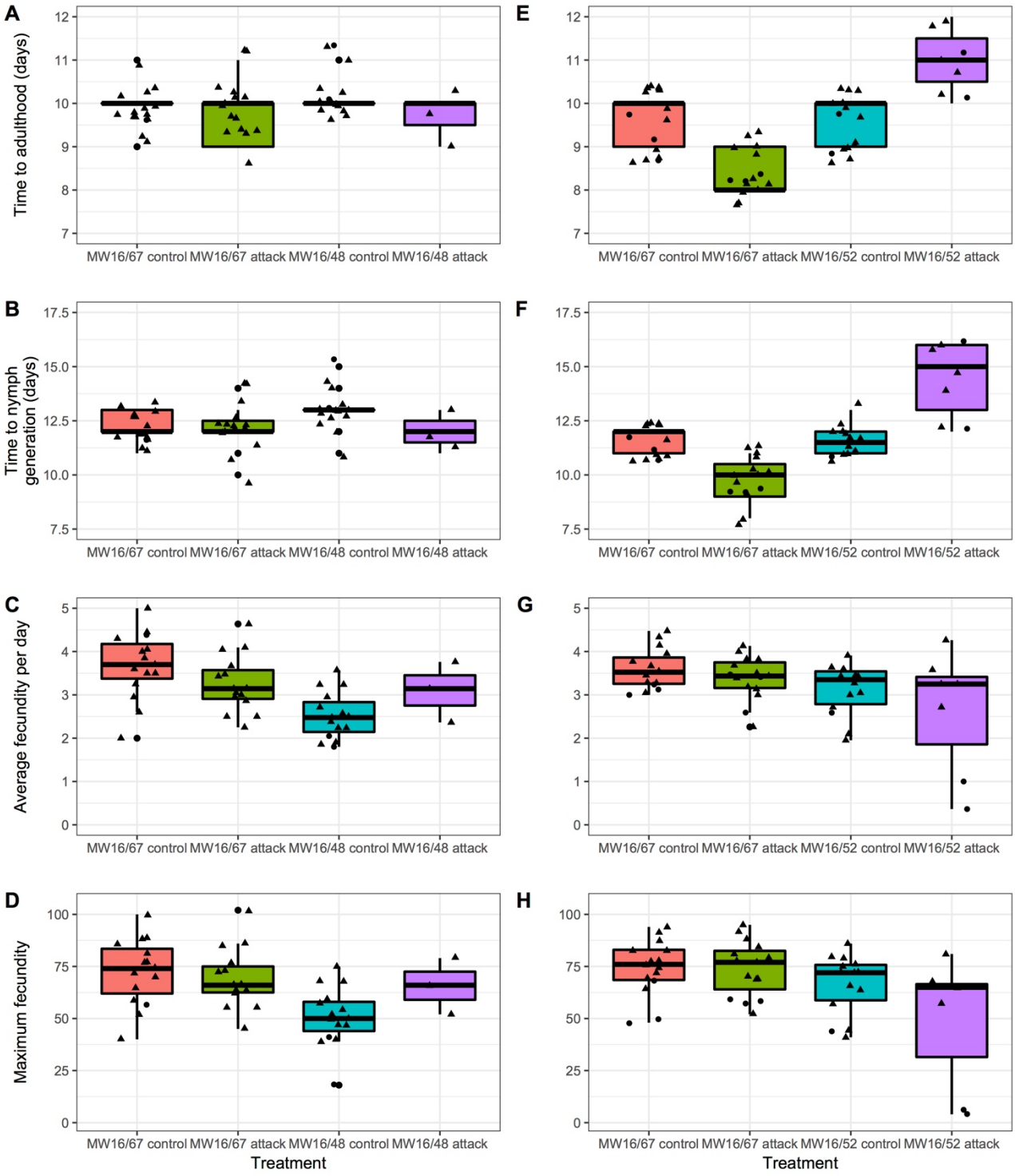
3.3.4 Parasitoid attack does not affect aphid development and fecundity in genotype 1

Aphid development to adulthood, first reproduction and adult fecundity were measured over two experiments, one comparing genotypes 1 and 3, and a second comparing genotypes 1 and 2. It was necessary to test whether genotype 1 aphids are affected equally over the two experiments. MANOVA results between genotype 1 in the two experiments was significantly different in both 'control' (Wilks = 0.675, F = 3.013 (1,28), $P < 0.05$) and 'attack' aphids (Wilks = 0.349, F = 11.655 (1,28), $P < 0.005$). Amongst the control aphids, ANOVA showed that days to nymph generation was significantly altered (F = 10.588 (1,28), $P < 0.005$), while in attacked aphids both days to adulthood (F = 46.411 (1,28), $P < 0.005$ and days to nymph generation (F = 39.444 (1,28), $P < 0.005$) were significantly different. Due to genotype 1 aphids showing different response behaviours, the two experiments were analysed separately.

In the comparison of genotypes 1 and 2, no parasitism occurred in genotype 1 (MW16/67) aphids, whereas 12 individuals were successfully parasitized in genotype 2 (MW16/48). There were no differences between the genotypes in their rate of development to adulthood or first reproduction (figure 7A,B) , although there was a trend towards shorter time to nymph production in attacked genotype 1 compared with unchallenged genotype 2 (figure 7B) (sup. table 8). Within each genotype, maximum fecundity and average fecundity per day did not vary in response to parasitoid attack, although fecundity was higher in attacked and unattacked genotype 1 aphids compared with unchallenged genotype 2 (figure 7C, D). Genotype 2 survival was strongly affected by parasitism, and only three individuals survived attack (Log rank test, $P < 0.01$). Genotype 1 survival was not affected by parasitoid attack, while genotype survival was reduced by attack (Log rank test, $P = 0.32$) (figure 8A).

In the comparison of genotypes 1 and 3, genotype 1 aphids again showed no parasitism by *A. ervi,* while 9 aphids were parasitzed in genotype 3 (MW16/52).

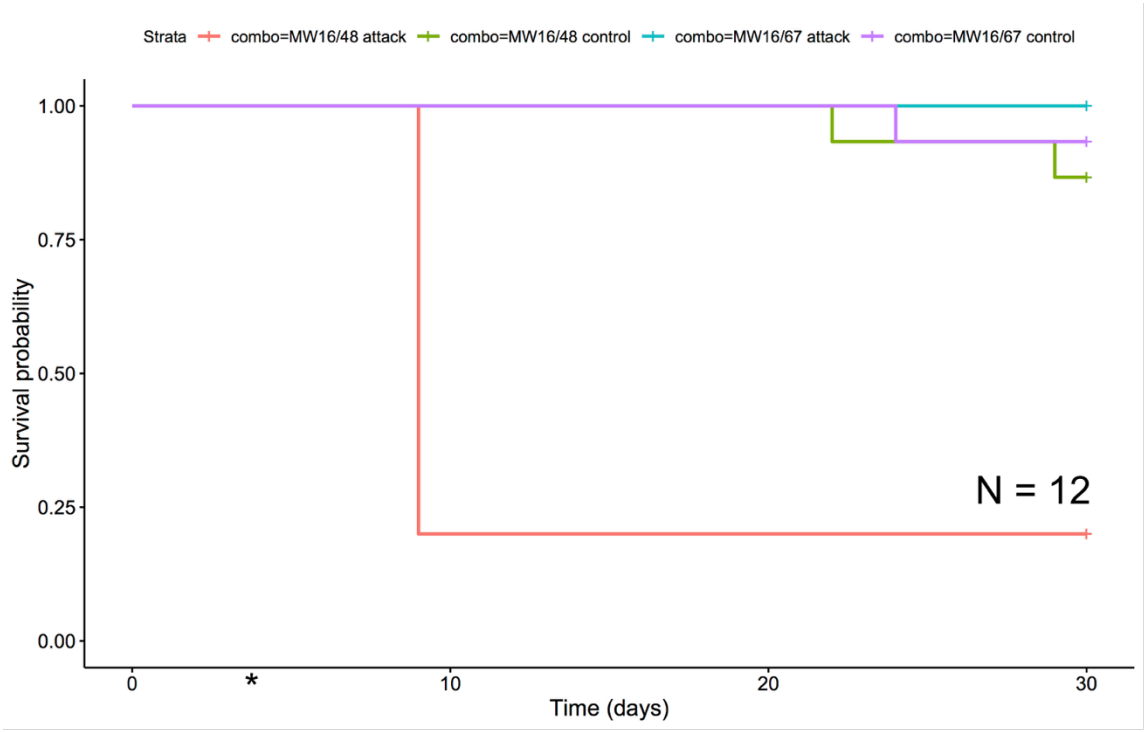
Attacked genotype 1 aphids showed significantly shorter time to reach adulthood and producing first offspring (figure 7E,F). Average max fecundity was not significantly affected by parasitoid attack in genotype 1 in comparison with genotype 3 (figure 7G) (sup. table 7). Challenged genotype 1 aphids showed increased total fecundity compared with unattacked and attacked genotype 3 aphids (figure 7H). In genotype 3 aphids, unattacked aphids showed higher average fecundity than attacked aphids. Finally, genotype 1 survival was not adversely affected by wasp oviposition (Log rank test, $P = 0.99$) while it was in genotype 3 (Log rank test, $P < 0.01$) (figure 8B).



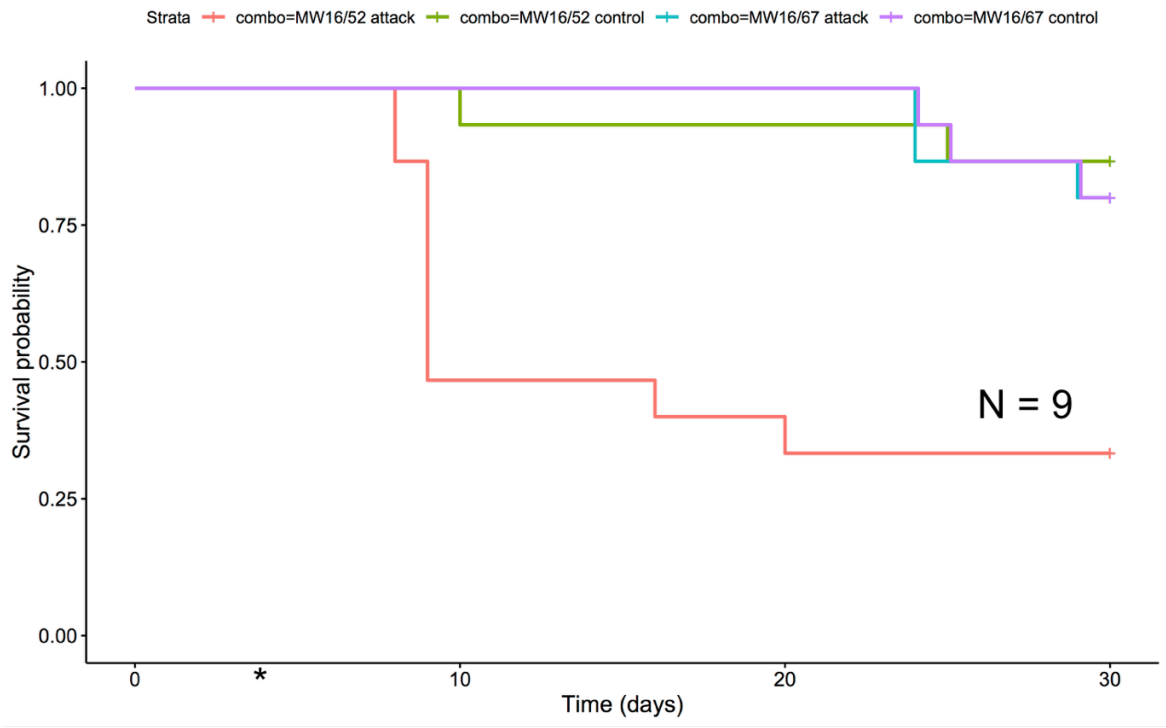
MW16/67 = genotype 1; MW16/48 = genotype 2; MW16/52 = genotype 3

Figure 7. *M. euphorbiae* development and fecundity in response to *A. ervi* attack compared with unattacked controls for genotypes that are resistant (genotype 1) and susceptible to parasitism (genotypes 2 and 3). A-D) MW16/67 compared to MW16/48. E-F) MW16/67 compared to MW16/52. Symbols illustrate individuals which survived 30 days (triangle) or died (circle) before 30 days.

A



B



MW16/67 = genotype 1; MW16/48 = genotype 2; MW16/52 = genotype 3

Figure 8. Survival probability of aphids in response to wasp attack for A) genotypes 1 and 2, and B) genotypes 1 and 3. N = number of aphids parasitized in susceptible lines. No parasitism occurred in resistant genotype 1 lines.

3.3.5 Genotype 1 potato aphids are more cold-tolerant than genotypes 2 and 3

GLMM analysis demonstrated significantly higher survival in adult genotype 1 aphids (10 days old) compared to genotype 3 (df = 341, z = -3.08, P < 0.01) and genotype 2 (df = 341, z = -3.54, P < 0.01) adult aphids (sup. table 9) (figure 9C). Genotype 1 adult aphids also showed significantly increased survival compared with 3-day old aphid nymphs of genotype 1 (df = 341, z = -41.5, P < 0.01) (figure 9A), but not with 7-day-old nymphs (figure 9B). At 10 days-old, cold exposure did not increase mortality of genotype 1 aphids compared to the warmer control conditions, while cold treatment increased adult mortality of genotypes 2 (df = 341, z = 6.46, P < 0.01) and 3 (df = 341, z = 2.88, P < 0.01). When assessed at the level of clonal line, the genotype 1 clonal lineage MW16/67 showed significantly higher survival in 10-day-old aphids compared to aphids of the genotype 2 lines MW16/48 (df = 323, z = -2.71, P < 0.01) and AK15/01 (df = 323, z= -3.25, P < 0.01), as well as the genotype 3 lines MW16/52 (df = 323, z = -3.05, P < 0.01) and MW17/04 (df = 323, z = -2.46, P < 0.05) (sup. table 10) (figure 9C). Significant differences in survival were not observed between the two genotype 1 lines MW16/67 and MW17/28 (df = 323, z = -1.81, P = 0.07), suggesting cold tolerance was associated with genotype rather than clonal differences.

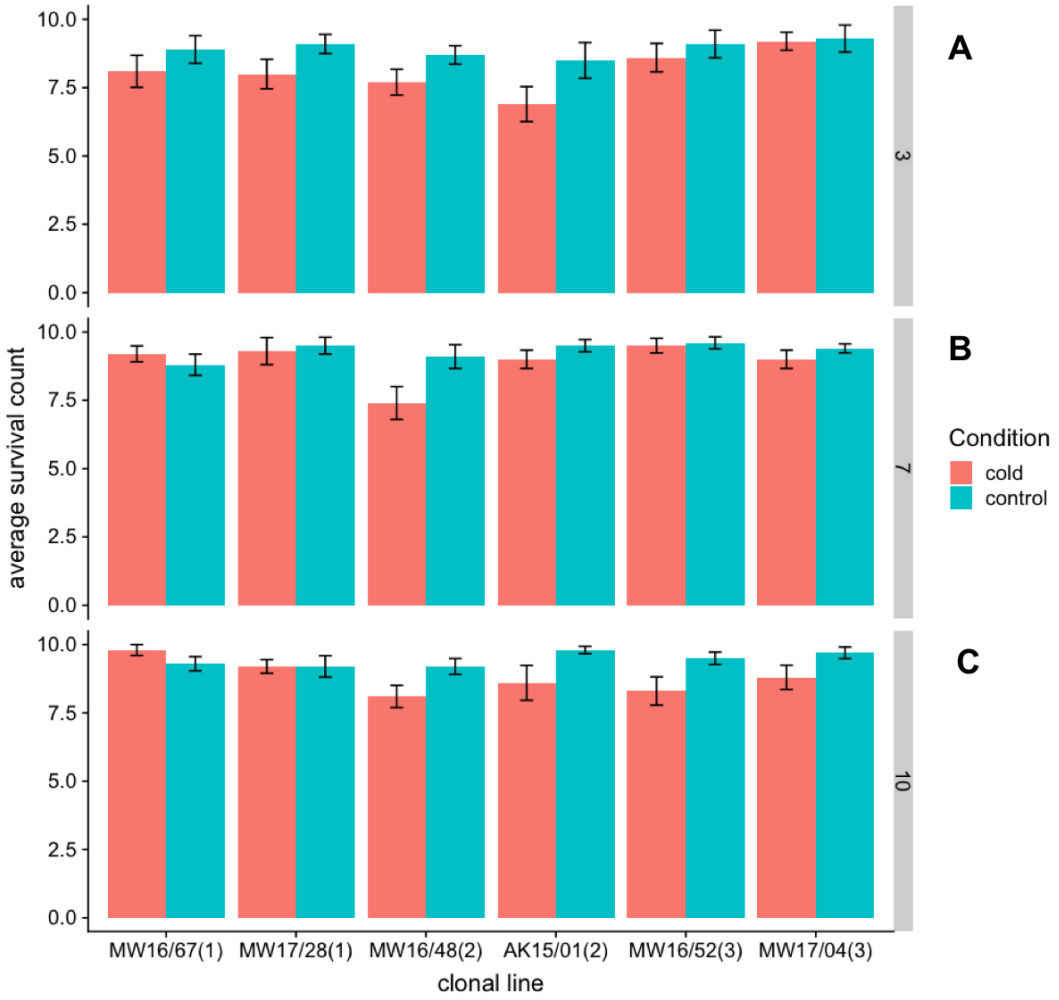


Figure 9. Aphid survival in response to sub-zero temperatures compared to controls maintained at 10 °C. Survival was measured 24 hours after cold exposure in nymphs/adult aphids at 3, 7 and 10 days-old (A, B, and C respectively). Averages generated from 10 replicates each consisting of 10 individuals. Numbers in parentheses indicate genotype identity determined through microsatellite analysis. Error bars indicate standard error of the mean.

3.4 Discussion

3.4.1 Genotype 1 advantageous traits do not confer high abundance in M. euphorbiae populations

The work undertaken here is the first to define the population structure of the Potato aphid *M. euphorbiae*, identifying the most and least prevalent genotypes. Specifically, the parasitoid resistant genotype 1 is surprisingly low in the population compared to highly abundant genotypes 2 and 3. Potato aphids belonging to genotype 1 were at a low frequency in wild populations associated with potato crops at commercial and garden/allotment sites in the three years of this study. This suggests a trade-off to parasitoid resistance that limits their population size. Low genotype 1 numbers may also indicate other factors are more important in maintaining the observed Potato aphid distribution. While not specifically studied here, Potato aphid genotypes may be differentially affected by the changes in phloem sap that occur over summer, resulting a crash in aphid numbers (Karley *et al.,* 2003). Other drivers here could be insecticides (in commercial farm settings) or the presence of other common predatory enemies, such as ladybirds, hoverflies and entomopathogenic fungi. Natural enemy activity can be effective at regulating potato aphid populations (Karley *et al.,* 2003). If parasitoids accounted for a greater proportion of this reduction, perhaps genotype 1 would become more prevalent throughout wild populations.

The most commonly observed genotype was genotype 2, followed by genotype 3. Variants of these two genotypes may also exist, indicated by the variability in size at the *Me5* marker site in genotype 2. Whether these variants translate into a specific phenotype is unclear. Genotype 2 aphids have been shown to cope better with poor quality host plant species (Karley *et al.,* 2017), suggesting they might be able to exploit a broader host range than other genotypes. Although the extent to which this characteristic affects their host range is undetermined, it may allow for easier dispersal, establishment and persistence of genotype 2 aphids. Genotype 3 was not found on potato in commercial fields, but was significantly more abundant in smaller allotments style settings. Genotype 3 aphids have been

shown to perform poorly on several *Solanum* species (Karley *et al.* 2017), which might influence its ability to establish on potato crops. In small allotments, plant species diversity is higher than that of a commercial field containing a crop monoculture and might provide a broader range of hosts to where genotype 3 aphids can persist. Microsatellite analysis with adegenet and STRUCTURE suggested little genetic similarity between the most frequent genotypes 2 and 3, and these were also the most distinct genotypes.

The more abundant microsatellite marker clusters of *M. euphorbiae* genotypes were consistent between years. This agrees with *M. euphorbiae* populations existing mainly as anholocyclic outside North America (Raboudi *et al.,* 2011). Some individual collections were unplaced within other clonal groups and may represent a handful of previously uncharacterized clones or perhaps even holocyclic variants. However, it is also likely that low quality of DNA extractions/re-calibration of instruments (i.e. 3730 DNA Analyzer) can also effect results. Genotype 6 appears more prevalent in the Merseyside area for reasons unknown. Although there was some overlap in microsatellite analysis between genotypes 1 and 6, similarities in parasitoid resistance between them are not as obvious. A similar possible relation is also observed within genotypes 2 and 7. While microsatellite data is reliable to define Potato aphid genotypes however, it does not provide enough resolution to distinguish clearly between all clones. Genotyping clustering is further improved through the use of whole genome sequencing and variant calling, and is further discussed in chapter 4.

3.4.2 Parasitoid resistant clones show no detrimental effects in response to attack

Genotype 1 potato aphids are highly parasitoid resistant. While a caveat here may be the small study size, wasp attack did not appear to have any detrimental effect on aphid development and fecundity. Instead, time taken to reach adulthood and first reproduction appeared to decrease in genotype 1 in response to challenge by *A. ervi*, possibly in-keeping with fitness compensation as observed in other aphid species in response to infection/predation (Altincick *et al.,* 2008; Vorburger *et al.,*

2013). This more rapid development did not translate to an increase in maximum fecundity or average fecundity per day, which were comparable with the control unattacked aphids. The mounting of an immune response to any pathogen can lead to detrimental effects on other life history traits of any organism. For example, Barribeau *et al.* (2014) demonstrated that inoculation of aphids with heat-killed Gram-negative and Gram-positive bacteria, as well as the aphid specific fungal pathogen *Zoophthora occidentalis*, had negative effects on aphid longevity and total number of offspring.

Upon parasitism challenge, there was little-to-no negative effect in genotype 1 Potato aphids in terms of development to adulthood or nymph generation. This is especially apparent when comparing with genotype 3, where aphid development and time taken to generate offspring were significantly slowed following parasitoid attack compared with unattacked aphids, leading to reduced survival probability in genotype 3 amongst aphids surviving parasitoid attack. One of the main constituents of *A. ervi* venom that aid in successful parasitism of a host are suggested to cause castration, where apoptosis is induced within aphid ovarioles through the increase of reactive oxygen species (Digilio *et al.,* 2000; Colinet *et al*., 2014). If potato aphid fecundity is not significantly affected by castration, it might indicate physiological mechanisms which contributes to *M. euphorbiae* resistance to *A. ervi* parasitism.

3.4.3 Genotype 1 shows greater cold tolerance than genotypes 2 and 3

Parasitoid resistant costs have been previously observed in *M. persicae*, where behavioural responses of insecticide-resistant genotypes to wasp attack result in higher levels of mummification compared with susceptible genotypes (Foster *et al.,* 2007; Foster *et al.,* 2010). Similarly, Pea aphids showing symbiont-conferred resistance exhibit lower fecundity compared with uninfected (susceptible) aphids (Gwynn *et al.,* 2005). To investigate aspects of trade-offs with parasitoid resistance, cold tolerance was assessed in genotypes 1,2 and 3 at varying ages, where cold-tolerance and overwintering ability may affect genotype distribution.

Variation in cold tolerance between aphid clones of the same species has been observed previously (Vorburger, 2004). With the high frequency of genotype 2 *M. euphorbiae*, increased cold tolerance could allow greater survival over winter, leading to higher prevalence than other genotypes earlier in the year, with the opposite potentially explaining low genotype 1 numbers. However, genotype 2 was the least cold resistant of the three genotypes tested, and suggests that overwinter survival is not the main factor explaining the prevalence of this genotype in potato aphid populations. Young nymphs (instars 1 and 2) are known to possess greater tolerance to cold exposure compared with adult aphids (Clough *et al.,* 1990). However, in comparison to genotypes 2 and 3, the parasitoid resistant genotype 1 aphids showed increased cold-hardiness as they aged. Further work could be carried out to examine the effects of cold exposure on fecundity and long-term survival.

3.4.4 Conclusion

This study has demonstrated that genotype 1 *M. euphorbiae* are possibly much hardier than other genotypes, showing considerable co-variance in survivability in response to multiple different stresses. However, how these effects shape the overall genetic structure of the potato aphid remains unclear, and requires further investigation. The lower frequency of genotype 1 indicates other variables (e.g. insecticides) are responsible for driving the observed distribution. Further work could include studies that encompass larger swathes of the UK rather than focussed studies around Tayside and, to a lesser extent, Merseyside, and make further links between overall distribution of possible host plants and differences in control/biocontrol usage across commercial settings. Host range especially could play a crucial role, as seen through distribution of genotype 3 solely on allotment settings and the ability of genotype 2 survive on poorer quality crops (Karley *et al.,* 2018). With regard to the causative effects of parastoid resistance and cold tolerance in genotype 1, these will be further discussed in later chapters through comparative genomic approaches.

3.5 Supplementary tables

Supplementary table 1. Primers used for amplification of microsatellite markers used in aphid genotyping. All primers outlined by Raboudi, *et al.,* 2005.

Microsatellite	Primer	Sequence (5'- 3')
Me1	Me1F	[6-FAM]- TTCGCGAAAAACTTTATGACC
	Me1R	TCGCTGCGTTCCTATACTACC
Me5	Me5F	[6-FAM]- GCAAATATTAAGGGTACAG
	Me5R	CCAATTAAAACAACTTCGTGG
Me7	Me7F	[6-FAM]- TTAAGTCACTGCCGGTTCG
	Me7R	ATTAGCTCGAGCTCGTAC
Me9	Me9F	[6-FAM]- AGCGAAACCTCCCCTAATAG
	Me9R	GCACAAATAAGCTCGAGTGC
Me10	Me10F	[6-FAM]- TCGCTGCGAGACTCGTATTG
	Me10R	GACGACGACGTGTACAATG
Me11	Me11F	[6-FAM]- CGTTTTCTACCCAAAGGAGG
	Me11R	ATTGTCCGTATACCACGACG
Me13	Me13F	[6-FAM]- GAACTCACTCAGACTCGTGTGG
	Me13R	CAGCCGGAATACCAAGAGC

Supplementary table 2. Reagents for microsatellite PCR.

Reagent	µl per reaction
5× clear GoTaq® reaction buffer	5.0
dNTPs	0.5
Nuclease-free water	15.875
GoTaq® DNA polymerase	0.125
Forward primer ± 5'[6-FAM] (10µM)	1.25
Reverse primer (10µM)	1.25
DNA template	1.0
Total	**25.0**

Supplementary table 3. Conditions for microsatellite PCR. An annealing temperature of 54°C was used for *Me1, Me5, Me7* and *Me9* primer sets. 62°C was used for *Me10, Me11* and *Me13* primer sets.

Time	Temperature	Repeat	Action
2 minutes	95°C	N/A	Initial denaturation
1 minute	95°C	40 cycles	Denaturation
1 minute	54/62°C		Annealing
1 minute	72°C		Extension
5 minutes	72°C	N/A	Final extension

Supplementary table 4. Primers used for amplification of bacterial sequences used for detecting symbiont infection.

Gene	Primer	Sequence (5'- 3')	Source
16S rRNA	16F27	AGAGGTTTGATCMTGGCTCAG	Lane, 1991
	1494R	GCTCTAGAGCGGYTACCTTGTTACGACTT	Lane, 1991
16S-23S rRNA	10F	AGTTTGATCATGGCTCAGATTG	Sandström *et al*., 2001
	480R	CACGGTACTGGTTCACTATCGGTC	Sandström *et al*., 2001
S. symbiotica 16S	16SA1	AGAGGTTGATCMTGGCTCAG	Fukatsu & Nikoh, 2000
	PASScmp	GCAATGTCTTATTAACACAT	Fukatsu *et al*., 2000
H. defensa 16S	PABSF	AGCGCAGTTTACTGAGTTCA	Darby & Douglas, 2003
	16SB1	TACGGYTACCTTGTTACGACTT	Fukatsu & Nikoh, 2000
R. insecticola 16S	U99F	ATCGGGGAGTAGCTTGCTAC	Sandström *et al*., 2001
	16SB1	TACGGYTACCTTGTTACGACTT	Fukatsu & Nikoh, 2000

Supplementary table 5. Reagents for symbiont PCR.

Reagent	**µl per reaction**
5× clear GoTaq® reaction buffer	5.0
dNTPs	0.5
Nuclease-free water	16.2
GoTaq® DNA polymerase	0.2
Hha 1	0.1
Forward primer (10µM)	1.0
Reverse primer (10µM)	1.0
DNA template	1.0
Total	**25.0**

Supplementary table 6. Conditions for symbiont PCR.

Time	**Temperature**	**Repeat**	**Action**
2 minutes	95°C	N/A	Initial denaturation
30 seconds	95°C	35 cycles	Denaturation
30 seconds	55°C		Annealing
3 minutes	72°C		Extension
7 minutes	72°C	N/A	Final extension

Supplementary table 7. Tukey's post-hoc comparisons for aphid development and fecundity in genotypes 1 and 3 in response to parasitism. '*' denote significant values, while '+' denotes approaching significance.

			95 % CI			
Factor	***Comparison***	***Mean difference***	***lower***	***upper***	***Sig.***	
Time to adulthood	gen3 + gen3 wasp	-2.86	-4.05	-1.67	0.00	*
	gen1 wasp + gen3 wasp	-4.70	-5.87	-3.52	0.00	*
	gen1 + gen3 wasp	-2.90	-4.07	-1.72	0.00	*
	gen1 wasp + gen3	-1.84	-2.79	-0.88	0.00	*
	gen1 + gen3	-0.04	-0.99	0.92	1.00	
	gen1 + gen1 wasp	1.80	0.86	2.74	0.00	*
Time to nymph production	gen3 + gen3 wasp	-2.86	-4.05	-1.67	0.00	*
	gen1 wasp + gen3 wasp	-4.70	-5.87	-3.52	0.00	*
	gen1 + gen3 wasp	-2.90	-4.07	-1.72	0.00	*
	gen1 wasp + gen3	-1.84	-2.79	-0.88	0.00	*
	gen1 + gen3	-0.04	-0.99	0.92	1.00	
	gen1 + gen1 wasp	1.80	0.86	2.74	0.00	*
Total Fecundity	gen3 + gen3 wasp	16.86	-4.05	37.76	0.15	
	gen1 wasp + gen3 wasp	24.44	3.77	45.11	0.01	*
	gen1 + gen3 wasp	24.84	4.17	45.51	0.01	*
	gen1 wasp + gen3	7.58	-9.20	24.36	0.63	
	gen1 + gen3	7.98	-8.80	24.76	0.59	
	gen1 + gen1 wasp	0.40	-16.09	16.89	1.00	
Average fecundity per day	gen3 + gen3 wasp	0.49	-0.38	1.36	0.44	
	gen1 wasp + gen3 wasp	0.76	-0.10	1.61	0.10	
	gen1 + gen3 wasp	0.96	0.10	1.82	0.02	*
	gen1 wasp + gen3	0.27	-0.43	0.96	0.73	
	gen1 + gen3	0.47	-0.23	1.16	0.29	
	gen1 + gen1 wasp	0.20	-0.48	0.88	0.86	

Supplementary table 8. Tukey's post-hoc comparisons for aphid development and fecundity of genotypes 1 and 2 in response to parasitoid attack. '*' denote significant values, while '+' denotes approaching significance. There was no significant differences or trends observed when comparing time to adulthood between genotypes 1 and 2, therefore Tukey's was not calculated on this occasion.

			95 % CI			
Factor	***Comparison***	***Mean difference***	***lower***	***upper***	***Sig.***	
Time to nymph production	gen2 + gen2 wasp	1.00	-0.54	2.54	0.32	
	gen1 wasp + gen2 wasp	0.13	-1.41	1.67	1.00	
	gen1 + gen2 wasp	0.27	-1.27	1.81	0.97	
	gen1 wasp + gen2	-0.87	-1.76	0.02	0.06	+
	gen1 + gen2	-0.73	-1.62	0.16	0.14	
	gen1 + gen1 wasp	0.13	-0.76	1.02	0.98	
Total Fecundity	gen2 + gen2 wasp	-14.67	-39.48	10.15	0.40	
	gen1 wasp + gen2 wasp	3.27	-21.55	28.08	0.98	
	gen1 + gen2 wasp	6.80	-18.01	31.61	0.88	
	gen1 wasp + gen2	17.93	3.61	32.26	0.01	*
	gen1 + gen2	21.47	7.14	35.79	0.00	*
	gen1 + gen1 wasp	3.53	-10.79	17.86	0.91	
Average fecundity per day	gen2 + gen2 wasp	-0.57	-1.69	0.55	0.53	
	gen1 wasp + gen2 wasp	0.16	-0.96	1.28	0.98	
	gen1 + gen2 wasp	0.59	-0.53	1.71	0.51	
	gen1 wasp + gen2	0.74	0.09	1.38	0.02	*
	gen1 + gen2	1.16	0.51	1.81	0.00	*
	gen1 + gen1 wasp	0.43	-0.22	1.07	0.31	

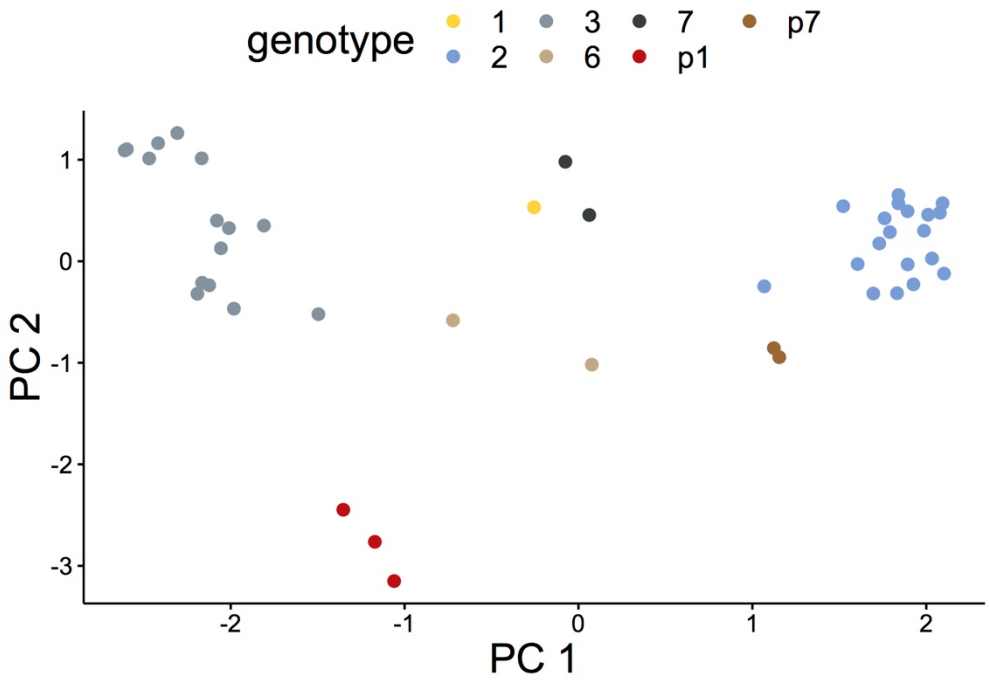
Supplementary table 9. GLMM analysis of aphid survival in response to cold at level of genotype. Baseline category were cold exposed genotype 1 at age 10 days-old.

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*	
(Intercept)	2.9893	0.367	8.146	3.78E-16	***
Condcontrol	-0.4711	0.4851	-0.971	0.33148	
Age3	-1.5407	0.3713	-4.15	3.33E-05	***
Age7	-0.4345	0.4218	-1.03	0.30294	
Genotype2	-1.3341	0.3772	-3.537	0.000405	***
Genotype3	-1.1793	0.3825	-3.083	0.002048	**
Condcontrol:Age3	1.2253	0.515	2.379	0.017344	*
Condcontrol:Age7	0.2984	0.5603	0.532	0.594381	
Condcontrol:Genotype2	1.7666	0.5649	3.127	0.001764	**
Condcontrol:Genotype3	1.8455	0.5899	3.129	0.001757	**
Age3:Genotype2	0.9035	0.4476	2.018	0.043564	*
Age7:Genotype2	0.3279	0.499	0.657	0.511045	
Age3:Genotype3	1.8605	0.4797	3.879	0.000105	***
Age7:Genotype3	1.1793	0.5396	2.186	0.028839	*
Condcontrol:Age3:Genotype2	-1.7184	0.6886	-2.496	0.012574	*
Condcontrol:Age7:Genotype2	-0.5498	0.7527	-0.73	0.465131	
Condcontrol:Age3:Genotype3	-2.2813	0.745	-3.062	0.002197	**
Condcontrol:Age7:Genotype3	-1.2769	0.8136	-1.569	0.116536	

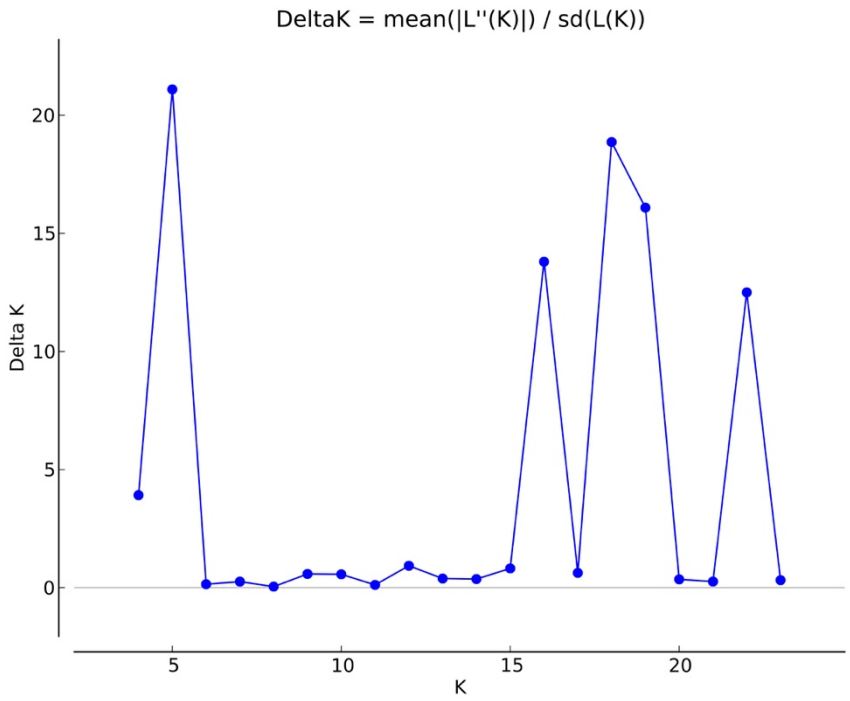
Supplementary table 10. GLMM analysis of aphid survival in response to cold at level of clonal line. Baseline category were cold exposed MW16/67 (genotype 1) at age 10 days-old.

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*	
(Intercept)	3.9407	0.7336	5.371	7.81E-08	***
Condcontrol	-1.348	0.8489	-1.588	0.1123	
Age3	-2.4587	0.7577	-3.245	0.00118	**
Age7	-1.4554	0.8029	-1.813	0.06989	.
LineageAK15/01(2)	-2.0884	0.7695	-2.714	0.00665	**
LineageMW16/48(2)	-2.4587	0.7578	-3.245	0.00118	**
LineageMW16/52(3)	-2.3211	0.7616	-3.048	0.00231	**
LineageMW17/04(3)	-1.9093	0.7771	-2.457	0.01401	*
LineageMW17/28(1)	-1.4554	0.803	-1.812	0.06992	.
Condcontrol:Age3	1.9622	0.9105	2.155	0.03116	*
Condcontrol:Age7	0.8604	0.9445	0.911	0.36233	
Condcontrol:LineageAK15/01(2)	3.3942	1.1195	3.032	0.00243	**
Condcontrol:LineageMW16/48(2)	2.3142	0.9288	2.492	0.01272	*
Condcontrol:LineageMW16/52(3)	2.6791	0.971	2.759	0.0058	**
Condcontrol:LineageMW17/04(3)	2.7992	1.0484	2.67	0.00758	**
Condcontrol:LineageMW17/28(1)	1.3109	0.9661	1.357	0.1748	
Age3:LineageAK15/01(2)	1.4263	0.84	1.698	0.08952	.
Age7:LineageAK15/01(2)	1.8413	0.9167	2.009	0.04458	*
Age3:LineageMW16/48(2)	2.213	0.8351	2.65	0.00805	**
Age7:LineageMW16/48(2)	1.0442	0.8737	1.195	0.23203	
Age3:LineageMW16/52(3)	2.6914	0.8542	3.151	0.00163	**
Age7:LineageMW16/52(3)	2.8261	0.963	2.935	0.00334	**
Age3:LineageMW17/04(3)	2.9126	0.8979	3.244	0.00118	**
Age7:LineageMW17/04(3)	1.6621	0.923	1.801	0.07173	.
Age3:LineageMW17/28(1)	1.3906	0.8797	1.581	0.11393	
Age7:LineageMW17/28(1)	1.6007	0.9671	1.655	0.09789	.
Condcontrol:Age3:LineageAK15/01(2)	-3.0888	1.2437	-2.483	0.01301	*
Condcontrol:Age7:LineageAK15/01(2)	-2.1941	1.3438	-1.633	0.10251	
Condcontrol:Age3:LineageMW16/48(2)	-2.2585	1.0844	-2.083	0.03727	*
Condcontrol:Age7:LineageMW16/48(2)	-0.5781	1.1261	-0.513	0.60769	
Condcontrol:Age3:LineageMW16/52(3)	-2.8262	1.1472	-2.464	0.01376	*
Condcontrol:Age7:LineageMW16/52(3)	-1.9974	1.2823	-1.558	0.1193	
Condcontrol:Age3:LineageMW17/04(3)	-3.306	1.2476	-2.65	0.00805	**
Condcontrol:Age7:LineageMW17/04(3)	-1.7922	1.2722	-1.409	0.15892	
Condcontrol:Age3:LineageMW17/28(1)	-1.023	1.134	-0.902	0.36702	
Condcontrol:Age7:LineageMW17/28(1)	-0.5032	1.2362	-0.407	0.68394	

3.6 Supplementary figures



Supplementary figure 1. PCA plot of 2016 *M. euphorbiae* genotypes based on seven microsatellite positions. All genotypes show little-to-no overlap with each other. As a result, only four loci were used in 2017 and 2018 using marker positions *Me1, Me5, Me9* and *Me10.* PC1 and PC2 explain 33.1% and 9.1% of the data.



Supplementary figure 2. Predicted value of genetic clusters (*K*) for *M. euphorbiae* based on the delta *K* method implemented in STRUCTURE HARVESTER (Earl & von Holdt, 2012). Briefly, delta K is calculated through measuring the highest likelihood of each value of *K*, followed by assessing the variance of likelihood values between replicates within each *K.*

3.7 Other electronic resources

chap1_R_comms.pdf
A list of R commands relevant to statistical analysis used in to analyse genotype distribution and phenotypical responses in *M. euphorbiae.*
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

4. Genome assembly of parasitoid resistant *Macrosiphum euphorbiae*

Abstract

Studying the genetic components of phenotypical variance between clonal genotypes in *Macrosiphum euphorbiae* requires a high-quality genome. Here, long-range information (10x linked-reads and HiC) has been utilized to scaffold a PacBio "gold" standard reference genome of genotype 1 *M. euphorbiae*, which presents near complete resistance to a predatory parasitoid wasp enemy. An initial Canu assembly contained multiple haplotypes, so Haplomerger2 was used to provide a haploid genome before scaffolding. Following the addition of linked-reads and HiC information, scaffold number decreased from 3,159 to 1,918, while N50 rose from 776 Kb to 57,542 Mb. Draft genome annotation yields 26,802 genes (94.5% complete BUSCOs), while repeat annotation demonstrates a high amount of repetitive sequence (38.06% of the genome). High gene counts and repeats are in line with *Acyrthosiphum pisum,* the closest relative aphid sequenced. *B. aphidicola* (*M. euphorbiae*) was assembled into a single chromosome with two plasmids (pLeu and pTrp), with a genome size of 656.5 Kb in total. The current gold standard genome of both the potato aphid and its endosymbiont will be used for variant identification between the other characterized genotypes of *M. euphorbiae*, as well as potentially evolutionary changes between other aphid species.

4.1 Introduction

4.1.1 The rise of insect genome sequencing

Insect molecular studies have been invaluable in elucidating genetic mechanisms behind phenotypical differences in individuals and populations. This is especially helpful within the agricultural sector and the study of crop pests, where individuals have developed resistance to control measures, such as chemical insecticides, and spread through insect populations. An example of this in aphids is the peach-potato aphid *Myzus persicae* Sulzer, where multiple genetic changes have arisen to provide resistance to a variety of pesticides, such as to pirimicarbs, pyrethroids and organophosphates, as reviewed in Bass *et al.* (2014). Detoxification of these insecticidal compounds is not restricted to these aphids however, with multiple types of resistance also observed in mosquito (Liu, 2015) and *Drosophila* species (Ffrench-Constant *et al*., 2004).

The rise of whole genome sequencing has led to many insect genomes being characterized, beginning with *Drosophila melanogaster* (Adams *et al.,* 2000). This has resulted in dedicated projects to aid in collaboration and comparative studies between multiple insect genomes, such as FlyBase (Thurmond *et al.,* 2018) and the i5K initiative (Robinson *et al.,* 2011). Genomic and transcriptomic studies for aphid species can be found at AphidBase (Legai *et al.,* 2010). Currently, genome assemblies for 7 aphid species are present, with all but the grape phylloxera aphid (*Daktulosphaira vitifoliae*) being publicly accessible. A genome sequence for the corn leaf aphid (*Rhopalosiphum maidis* Fitch) has also recently been made available (Chen *et al*., 2019) outside of AphidBase.

The sequence information for each aphid genome varies in quality due to different assembly methods (table 1). Principles of genome assembly dictate contig/scaffold number should be as low as possible, while N50 (where contigs of this length or greater make up half the overall genome assembly) should be a large as possible. Some have been generated from Illumina short reads only, with

and without mate pair libraries for scaffolding assembled contigs. The recent *R. maidis* genome differs in that it was generated using long PacBio reads and scaffolded using Hi-C, resulting in four near complete chromosomes (Chen *et al*., 2019). The advantage of Hi-C scaffolding is clearly observed in a recent release of the Pea aphid genome (*Acyrthosiphon pisum*), where the integration of long-range information increases contiguity, again providing highly complete chromosomes (Li *et al.,* 2019). During Hi-C preparation, chromatin is cross-linked *in vivo* to capture the chromosome 3D structure and sequences linked by proximity. The DNA is digested, leaving behind short crosslinked regions for Illumina sequencing, where regions between these ligated fragments can represent up to 100's of kilobases or megabases in length (Dudchenko *et al.,* 2017). The 10x Chromium platform provides a similar service for scaffolding and genome phasing (Mostovoy *et al.,* 2016; Seo *et al.,* 2016, Yeo *et al.,* 2017; Coombe *et al.,* 2018), although the technology relies on sequencing short fragments along long DNA fragments from 50-100 kb, where each of these fragments is clustered using a proprietary barcode system (10x Genomics, San Francisco, USA). Other advantages of the 10x platform are the small input of DNA required (as little as 1 ng), and the fact that it can be used for *de novo* genome assembly (Weisenfeld *et al.,* 2017), whereas mate-pair libraries and Hi-C facilitate contig scaffolding only. Finally, long-read sequencing and scaffolding can also be achieved through MinION nanopore technologies, providing real-time sequencing of DNA molecules. Compared with PacBio sequencing, data yield is often much lower and can have a higher error rate, such as incorrect base-calling over homo-polymer regions. However, the MinION has advantages in terms of being small and portable. Currently, MinION sequencing is suited perfectly to small bacterial genomes, where high coverage can easily be achieved (Karlsson *et al.,* 2015) but has also been useful in the scaffolding of eukaryotic Illumina-only genomes as well (Tan *et al.,* 2018).

Assembly	*Rhopalosiphum maidis + Hi-C*	*Acyrthosiphon pisum + Hi-C*	*Acyrthosiphon pisum*	*Aphis glycines*	*Diuraphis noxia*
num. contigs	220	21,425	23,452	3,209	5,636
Total length	326,023,155	540,926,991	541,490,866	308,063,388	395,057,805
GC (%)	27.69	29.76	29.76	27.27	29.07
N50	93,298,903	132,544,852	518,681	5,864,887	397,774
N's per 100 kbp	14.39	7,653	7,717	1,683	25,085

Assembly	*Myzus cerasi*	*Myzus persicae (Clone G006)*	*Myzus persicae (Clone O)*	*Rhopalosiphum padi*
num. contigs	49,286	4,022	13,407	15,587
Total length	405,711,039	347,304,760	354,698,803	319,422,546
GC (%)	29.87	30.03	30.19	27.77
N50	23,273	435,781	164,460	116,185
N's per 100 kbp	49.23	528.7	3,262	17.23

Table 1. Genome assembly statistics for currently available aphid genomes. All assemblies apart from *R. maidis* were generated using short reads. The use of long PacBio reads and Hi-C demonstrates increased levels of contiguity and chromosomal completeness, as indicated by much higher N50 values. See section 2.4 for genome data sources.

High quality genome assembly is important for any downstream analysis, especially when studying novel biology or generating complete gene annotations. Long-read sequencing and scaffolding can be key for spanning repetitive regions that are difficult to resolve with short reads alone (Lee *et al.,* 2016; Grau *et al.,* 2018). For example, assemblies for *A. pisum* contain approximately 40 Mb of gaps generated though the addition of mate-pair libraries on a short-read assembly. Without any filling of these gaps, it may inadvertently mask gene discovery as well as prevent the study of repetitive elements (e.g. transposable elements (TEs)). High contiguity can also help observe complex structural variants such as large scale inversions, and again is something that is difficult to infer from short-read data alone (Shao *et al.,* 2018). In aphid genomes, TE abundance can be the result of horizontal gene transfer (HGT) (Peccoud *et al.,* 2017), the movement of genetic material from one organism to the other. HGT may also carry genic content, with a common example in aphids is the integration of fungal carotenoid synthesis genes that result in changes to colour phenotype (Moran & Jarvik, 2010).

4.1.2 Gene annotation of current aphid genomes

Much like the differences in assembly quality between aphid genomes, reported gene annotations vary as well, with the smallest aphid gene set of 17,629 in *R. maidis* (Chen *et al*., 2019) compared with the largest set in *A. pisum* (36,939) (IAGC, 2010). The pea aphid genome is thought to have undergone extensive expansion in 2,459 gene families, which leads to an inflated gene count. However, this annotation was generated from a combination of *ab initio* predictors (no prior information, such as RNA-seq) and a cDNA set that provided only 11,089 gene models through the NCBI RefSeq pipeline. While duplication of these genes may be apparent, the number of genes that are functional could be much smaller. Thorpe *et al.* (2018) demonstrated through re-annotation of the pea aphid and several other aphid species that gene counts for all lie within a range of approximately 25,000 to 27,000 genes. Perhaps excessive gene numbers between aphids indicate the presence of pseudogenes, or is a result of different methods of gene annotation.

Illumina-generated RNA-seq and transcriptomics is currently the standard, but remains expensive, especially when running differential expression analyses. The main alternative is PacBio Iso-seq, which aims to characterise whole length mRNA molecules as well as discover possible gene iso-forms that can be harder to identify with an Illumina short-read approach. Iso-seq analysis is expensive as well. The recent release of RNA and cDNA sequencing using the Oxford Nanopore (ONT) MinION platform can overcome the challenge of cost, while also providing full length transcripts and even iso-form variants. While arguably more useful for bacterial genomes, effort has gone into eukaryotic gene annotation using this technology (Bayega *et al.,* 2018). Caveats include the potential for low output due to variability in flow-cell quality (e.g. pore availability), requiring the use of more MinION flow-cells and ultimately increasing the cost. However, it's accessibility and relative ease-of-use compared with other current methods demonstrate its potential.

4.1.3 Aims and objectives

Studying the genetic variation between *M. euphorbiae* genotypes to understand the mechanisms leading to their associated phenotypes, such as those that are resistant to parasitism by *A. ervi*, requires a draft genome and gene annotation. The following work will attempt to generate the following:

1. Generate PacBio genome assembly with comparisons to multiple genome assemblers. The best assembly will be further scaffolded using 10x linked-reads followed by integration with Hi-C scaffolding and orientation through the Dovetail Genomics HiRise pipeline, with the aim of generating a chromosome scale assembly (n = 5, Monti *et al.,* 2011).
2. Compare genome contiguity against other chromosome level assembly aphid assemblies, as well as comment genome evolution between these aphid species.
3. Generate draft gene annotation, comparing two current tools, Maker2 (Holt & Yandell, 2011) and BRAKER (Hoff *et al.,* 2015). RNA-seq and transcriptome assembly will be generated using Illumina data from a

previous study of *M. euphorbiae* (Teixeira *et al.,* 2018; NCBI accession: SRX339176) as well as attempting to use MinION transcripts generated through the cDNA direct kit.

4. identify aphid genes of interest, specifically those linked to immunity. Assess the potential for gene family expansion in immunity genes between other aphid species.
5. Generate a draft genome of the *Buchnera* symbiont, followed draft annotation with prokka.

This will provide a basis for the following chapter for exploring genetic differences in the potato aphid through read mapping and variant calling.

4.2 Methods

4.2.1 M. euphorbiae culture conditions for sequencing material

The parasitoid resistant aphid, genotype 1 line MW16/67, described in chapter 3 was used for *de novo* genome assembly. Six other clonal lines of genotype 1 (chapter 5, table 1) were also sequenced to provide population genomic data (see chapter 5). Unless otherwise stated, aphids were maintained under conditions outlined in section 2.1. For insect material used in high-molecular weight DNA extraction and Hi-C libraries, aphids were cultured at 16L:8D; 30 °C +/- 1; 70% RH, for 5 days prior to sampling. The higher culture temperature was used to reduce symbiont density (IAGC, 2010) and reduce symbiont relative abundance in sequencing data.

4.2.2 Genomic DNA extraction and preparation of short read Illumina libraries

Genomic DNA extraction was performed as outlined in section 2.2. For initial sequencing of six genotype 1 lines, library preparation was carried out by the Centre for Genomic Research (CGR) in the University of Liverpool. TruSeq PCR free libraries (2x150 bp) with a 550 bp insert were generated from six of the submitted gDNA samples. Libraries were sequenced on a single lane of the Illumina HiSeq 4000.

4.2.3 High molecular weight DNA preparation and Long-read library sequencing

The method described is adapted from the Quick protocol (Quick, 2018). Fifty adult aphids of each clonal line were divided between three 2 mL Eppendorf tubes. The samples were flash frozen in liquid nitrogen and homogenised using a pestle. One mL of tissue lysis buffer (10 mM Tris-Cl pH 8.0; 25 mM EDTA pH 8.0; 0.5% SDS) was added to the homogenised insect powder. The three aliquots were then pooled in a 50 mL Falcon tube with an extra 2 mL of lysis buffer. RNAse A (20 mg/mL) was added to a final concentration of 20 μg/mL, the tube slowly rotated end-over-end ten times and incubated at 37 °C for 1 hour. 25 μl Proteinase K (20

mg/ml) was then added and the solution incubated at 50 °C for 2 hours, with 10 more slow end-over-end rotations performed every hour.

The lysis solution was decanted into a 15 mL tube containing light-variety phase lock gel (Quantabio, Massachutsetts, USA), along with 5 mL of phenol solution. The tube was placed on a rotary shaker at 40 rpm for 10 minutes until a "milky" mixture had formed. The sample was centrifuged at 2,300 *g* for 12 minutes, after which the aqueous phase was poured into a new 15 mL tube containing phase lock gel. The gel is less dense than the organic phase of the phenol, resulting in a cleaner extraction and DNA which is less sheared compared to methods that require pipetting. 5 mL chloroform/isoamyl (25:1) was added and the sample placed back on the rotary shaker for 10 minutes at 40 rpm followed by centrifugation at 2300 *g* for 12 minutes. The aqueous phase was decanted into a 50 mL tube.

To precipitate the DNA, 15 mL of ice cold ethanol was added to the aqueous phase, along with 2 mL of sodium acetate (1M; pH 5.2) and mixed by careful inversion, rotating the tube gently end-over-end 10 times. The 'jelly'-like DNA precipitate was then "pulled" from the solution using a glass Pasteur pipette with a hooked end, made by briefly heating the end up in a Bunsen burner. DNA was washed by submerging in 70% ethanol until turning opaque, then was carefully removed from the pipette into a new 2 mL tube. DNA was centrifuged in 1 mL of 70% ethanol for 1 minute at 10,000 *g*. Ethanol was removed with a pipette without disturbing the DNA pellet. The pellet was then left to air-dry at 40 °C for 10 minutes to remove residual ethanol. DNA was re-suspended without agitation in 50 μl nuclease-free water and left to slowly dissolve overnight at 4 °C. DNA quality was assessed using 1 μl of DNA (diluted 1 in 10 in nuclease-free water), with purity measured using Nanodrop and quantity estimated using Qubit.

High molecular weight DNA library preparation and sequencing was performed by CGR, with resulting libraries sequenced on the PacBio Sequel using v1.2.1

chemistry. DNA was sheared to 20 Kb. DNA was sequenced over 10 SMRT cells, providing 60 Gb of data in total. The same high molecular weight DNA was used to prepare a single 10x chromium library and sequenced on a lane of the Illumina HiSeq 2500 (paired-end sequencing; 2x150 bp). Library preparation and sequencing for Hi-C libraries was performed by Dovetail genomics (California, USA).

4.2.4 RNA extraction for transcriptome sequencing

RNA was extracted from 200 individual aphids from clonal culture MW16/67 (genotype 1), consisting of a mixture of instars. Aphids were divided over four 2 mL Eppendorf tubes (50 aphids each). Aphids were flash frozen in liquid nitrogen and homogenised using a pestle. Cells were re-suspended in 500 μl TRI-reagent (Sigma-Aldrich, Missouri, USA) and left at room temperature for 5 minutes. RNA from the cell solution was isolated using the Direct-zol RNA MiniPrep kit (Zymo-research, California, USA) following the procedure outlined in the kit, as well as performing the optional DNase step. mRNA isolation was performed using Dyna-beads (Thermofisher, Massachusetts, USA) using the procedure outlined within the kit. Qubit and Nanodrop values were obtained throughout the process. As mRNA quantities were low after isolation with Dyna-beads, it was decided that integrity would be measured after cDNA synthesis to ensure as much material as possible was kept for MinION sequencing.

For transcriptome sequencing, 250 ng of purified mRNA was used as input for library preparation using Direct cDNA sequencing kit (SQK-DCS109) (Oxford Nanopore Technologies (ONT) Oxford, UK). The kit Direct cDNA requires no PCR to avoid introducing PCR bias. Library preparation was performed as outlined in the accompanying protocol, with a minor change. After reverse strand synthesis and adapter ligation, five separate libraries were pooled to generate as much sequencing material as possible. Pooled libraries were sequenced using two FLO-MIN106 version flow cells and MinKnow (v18.03.1). Base calling was performed

using Albacore (v2.3.1) (Wick *et al*., 2019), with reads that surpass ONT read quality threshold used in gene prediction.

4.2.5 Genome assembly of M. euphorbiae

Three long-read genome assemblers were compared: Canu (v1.4) (Koren *et al.,* 2017), HGAP (v3) (Chin *et al.,* 2013), and MaSuRCa (v3.2.8) (Zimin *et al.,* 2017). All were run using default parameters and were provided with a genome estimate size of 530 Mb (Wenger *et al.,* 2017). The MaSuRCa assembly was also provided with short Illumina reads generated from the same clonal line used for long-read sequencing. GenomeScope (v1) was used to predict genome size using the short-read Illumina library for MW16/67 (Vurture *et al.,* 2017).

4.2.6 Genome assembly polishing using arrow

As noisy PacBio reads can lead to errors in genome assembly, genome polishing was performed using the arrow algorithm on PacBio only assemblies, implemented through the smrtlink portal (v4.0.0). HGAP includes arrow polishing post-assembly, and was therefore not further used. The MaSuRCA assembly was also not further polished due to the assembly including short accurate Illumina reads. Three iterations of arrow were performed to polish the Canu assembly prior to any downstream analysis. For each iteration, the Canu reference was uploaded to the smrtlink portal and provided with the all PacBio reads used in the initial assembly.

4.2.7 Assessment of genome contiguity and gene completeness

Assembly statistics were generated using Quast (v5.0.0) (Mikheenko *et al.,* 2018). Gene content was assessed using BUSCO (v3) (Simão *et al.,* 2015; Waterhouse *et al.,* 2017). BUSCO searches for single copy orthologues present within a selected taxonomic group and allows inference of genome completeness. BUSCO was used

with the arthropod_odb9 database and used with 'geno', 'tran' or 'prot' mode for assessing genome, transcriptome or protein sets respectively.

4.2.8 Removing diploid sequence through Haplomerger2

Where duplication of arthropod orthologous genes was high, assemblies for *M. euphorbiae* were subject to HaploMerger2 (v20161205) (Huang *et al.,* 2017), a tool designed to provide a haploid assembly from a diploid sequence through self-alignment with LASTZ. Haplomerger2 also requires the use of a custom script (lastz_D_wrapper.pl) to generate a score matrix at 95% identity (Huang *et al.,* 2017). Haplomerger2 was used until genome size was reached of approximately 530 Mb based on Wenger *et al.* (2017) and GenomeScope analysis.

4.2.9 Scaffolding and contig sorting into chromosomal groups

Contig orientation and scaffolding was performed first with 10x linked reads using arcs (v1.0.1) (Yeo *et al*., 2017). 10x GemCode barcodes were manually appended to fastq headers and mapped using bwa mem (v0.7.5a) (Li & Durbin, 2009) and the resulting bam file provided to arcs. Further scaffolding with Hi-C data was performed by Dovetail genomics using their HiRise assembly platform (California, USA). Hi-C contact maps were generated using Juicer (v1.5) and visualised using Juicebox (v1.8.9) (Durand *et al.,* 2016). Progressive scaffolding was visualised using the distribution of arthropod BUSCOs across the Canu assembly with and without the incorporation of 10x and Hi-C data, where number of BUSCOs was plotted against sorted scaffolds from largest to smallest. Steeper curves indicate more gene content in fewer scaffolds. Plots were generated using R (v3.2.0) (R core team, 2015) and ggplot2 (v3.1.0) (Wickham, 2016). (see https://github.com/hlmwhite/scaffolding_assessment_using_BUSCO).

4.2.10 Visualising symbionts and assembly contamination with Blobtools

Blobtools (v0.9.19) (Laetsch & Blaxter, 2017) requires three inputs: a coverage file of mapped reads, a hits file (BLAST output) and a genome assembly. PacBio reads were mapped to the scaffolded Canu reference with blasr (v1.3.1.121193) (Chaison & Tesler, 2012), and the resulting SAM file provided to map2cov (Blobtools script). MegaBLAST (v2.2.26) (Chen *et al.,* 2015) was used to assign taxonomic IDs to scaffolds.

4.2.11 Data sources for comparative analyses

The following analyses use a range of arthropod and bacterial genomic, transcriptomic and proteomic data sets. All external data used is summarised in section 2.4.

4.2.12 Genome synteny between other aphids

The release of the *A. pisum* genome integrated with Hi-C (Li *et al.,* 2019) as well as the recent release of the chromosome scale *R. maidis* genome provides a useful tool for assessing synteny, conserved regions and genome evolution between these species and *M. euphorbiae*. Synteny was inferred from the distribution of shared complete single copy orthologues (see section 4.2.7). Resulting '.tsv' files generated through BUSCO were provided to a custom R script (available at https://github.com/hlmwhite/synteny_from_BUSCOs), generating a plot showing shared BUSCO genes between chromosomes of two genomes of interest. Plots were generated in R (R core team, 2015) and requires the R package BioCircos (Cui *et al.,* 2016).

Synteny was also observed through whole genome alignment. Genomes were aligned to each other using Minimap2 (v2.2-r424-dirty) (Li, 2018) with the options "-PD -x asm5" enabled. Resulting '.paf' files from Minimap2 were provided to the

online version of D-GENIES (Cababettes & Klopp, 2018). The options “Sort contigs” and “Hide noise” were enabled.

4.2.13 Repeatmasking and repeat element annotation

A species-specific repeat library was generated using RepeatModeler (v1.0s.11) (Smit & Hubley, 2008) with the arcs/HiRise scaffolded Canu assembly. RepeatMasker (v4.0.7) (Smit *et al*., 2013) was used to identify repeats and transposable elements within the scaffolded Canu assembly. Repeat elements for three other aphid species (*A. pisum, R.* maidis and *M. persicae* (G006)) were performed in this manner for comparison purposes. The ‘.out’ file was manipulated to generate a bed file to soft-mask the genome when required using bedtools makefastaFromBed (Quinlan & Hall, 2010), soft-masking low-complexity and simple repeats.

4.2.14 Transcriptome assembly of MinION and Illumina reads

Pinfish (v1) was used to generate reference-based transcriptome assemblies from MinION data (Nanorporetech, 2019). Reads are clustered on similarity, where cluster sizes below a specified cut-off are removed. The higher the cut-off, fewer reads are kept, while confidence in correct transcripts increases. Three cluster sizes were assessed: 2, 5 and 10. Trinity (v.2.7.0) (Grabherr *et al.,* 2011) was used to re-assemble Illumina transcriptomic reads from a previous study (Teixeira *et al.,* 2018; NCBI Sequence Read Archive accession SRX339176).

4.2.15 Draft genome annotation using Maker2

Maker2 (Holt & Yandell, 2011) is an iterative gene predictor for eukaryotic organisms, able to make use of a range of evidence types. Maker2 was provided with protein sets for *A. pisum, M. persicae, D. noxia, A. glycines* and *R. padi*, with protein sets clustered using CD-hit (v4.6) (Li & Godzik, 2006) to remove redundant proteins between species, as well as the *A. pisum* AUGUSTUS species model and a Genemark-ES file of gene predictions. Maker2 was also provided with one of two

assembled transcript sets: all assembled transcripts from Trinity and Pinfish; or full length transcripts from Trinity and Pinfish as determined through the Trinity accompanying script 'analyze_blastPlus_topHit_coverage.pl', where the length of assembled transcripts are compared to their closest BLASTx match in the Swissprot database. The first iteration of Maker2 was run with "est2genome=1" (infers gene predictions directly from assembled transcript alignments) and "protein2genome=1" (infers gene predictions directly from protein alignments) enabled, after which a SNAP gene model was generated (v2006-07-28) (Korf, 2004). Maker2 was executed twice more, being provided with SNAP models generated from previous iterations, but with "est2genome=0" and "protein2genome=0" turned off. A final iteration of Maker2 was performed with "keep_preds=1" enabled. Resulting genes were filtered on Annotation Edit Distance (AED) using the Maker2 script quality_filter.pl (AED < 1) and annotated with pfam domains using InterProScan (v5.30-69) (Jones *et al.,* 2014).

4.2.16 Draft genome annotation using BRAKER

BRAKER (v2) (Hoff *et al.,* 2015) was provided with mapped transcriptomic reads to a softmasked reference genome, where identified repeat structures through RepeatMasker are left in lower case. Illumina reads were mapped with STAR (v2.5.2a_modified) (Dobin *et al.,* 2013), while MinION reads were mapped using Minimap2 (v2.2-r424-dirty) (Li, 2018) using splice aware settings ('-ax splice'). BUSCO was used to assess completeness of all gene annotation methods. InterProScan was used to annotate proteins with GO terms and pfam domains.

4.2.17 Filtering predicted gene sets on expression

Predicted genes were filtered on mapped transcriptomic evidence to provide a set of expressed genes. A GTF file generated through BRAKER was provided to featureCounts (v1.6.3) (Lioa *et al.,* 2014), along with the reference *M. euphorbiae* genome and bam files for mapped Illumina and ONT transcriptomic reads. All genes with at least one transcript hit were taken to generate a final gene set for

M. euphorbiae and subsequently used in further analysis. BUSCO was used to assess transcriptome completeness.

4.2.18 Identification of immune function genes in M. euphorbiae

Genes with immunity-related functions were identified for further analysis into parasitoid resistance (see chapter 5). Gerardo *et al.* (2010) identified a list of 106 Pea aphid genes linked to immunity, belonging to one of one of 16 functional groups or pathways. Two clip-domain serine proteases (serpin) identified by Ma *et al.* (2019) were also included in this list (*AP_SPLP* and *Ap_VP*), both of which are thought to be involved in induction of phenoloxidase cascade and melanisation activity. Immune protein orthologues were identified using Orthofinder (v2.2.7) (Emms & Kelly, 2018) using default settings. Orthofinder was provided with protein sets for *A. pisum, M. euphorbiae, D. noxia, A. glycines* and *M. persiae* (O clone). Immune functioning proteins predicted in *M. euphorbiae* were confirmed through BLASTx ("-evalue 1e-40" enabled) against *A. pisum* transcripts (v2.0) available from AphidBase.

4.2.19 Genome assembly for the Potato aphid obligate symbiont, Buchnera aphidicola

The endosymbiont, *B. aphidicola,* was isolated from a short-read DISCOVAR DeNovo (vr52488) (Broad Institute, 2015) assembly of *M. euphorbiae* genotype 1 clonal line AA09/04. DISCOVAR DeNovo was run with default settings along with a predicted genome size of 500 Mb, where *Buchnera* was assembled into a single contig. *B. aphidicola* and its two plasmids (pTrp and pLeu) were identified via BLAST (v2.6.0+) (Camacho *et al.,* 2009) against a *Buchnera* reference from *A. pisum* (ncbi: GCA_000009605.1). As *Buchnera* was isolated from a different aphid clonal line used in aphid genome assembly, it was necessary to confirm the same *Buchnera* genotype between aphid clones. Short reads from MW16/67 were provided to Snippy (v4.3.6) (Seemann, 2015) to call variants against the *Buchnera* reference from AA09/04. Snippy identified two single nucleotide polymorphisms

(SNPs) and no other variants, therefore base changes were manually made at these positions.

4.2.20 Gene prediction and pseudogene identification for B. aphidicola

Gene prediction for endosymbiont assemblies was performed using Prokka (v1.12) (Seemann, 2014) with the options "--rfam" (enables searches for non-coding RNAs) and "--compliant" (forces GenBank compliance) enabled. Psuedogenes were predicted using pseudofinder and the settings "--length_pseudo 0.40" and "--shared_hits 0.60" to remove potential false-positives in pseudogene identification (Syberg-Olsen & Husnik, 2018). Gene prediction was performed for *B. aphidicola* generated through this study, as well as 15 other *Buchnera* genomes and *Escherichia coli* K-12, all available from NCBI. Pseudogene annotation was performed for three *Buchnera* strains: *B. aphidcola* from *A. pisum* (GCA_000009605.1), *B. aphidcola* from *M. euphorbiae* (GCA_005237295.1) and *B. aphidicola* from *M. euphorbiae* generated through this study.

It was necessary to identify sequence variation between *Buchnera* strains as well as identify amino acid synthesis genes, due to the vital mutualism observed between *Buchnera* and its host (Van Ham *et al.,* 2003), and the potential to cause phenotypic variation which will be further discussed in chapter 5. Comparisons were made using the Artemis Comparison Tool (ACT) within Artemis (v18.0.2) (Carver *et al.,* 2005). The *Buchnera* assembly generated here was manually reordered so that sequence began with the uridine modification enzyme encoding *mnmG,* as with *Buchnera* strains available through NCBI. BLAST comparisons (BLASTn with "-evalue 1" and "-m 8" enabled) were generated between the *Buchnera* assembly generated here and genomes for *Buchnera* strains belonging to *A. pisum* and *M. euphorbiae* available from NCBI. ACT was provided with BLASTn results as well as gene features identified through Prokka and pseudofinder analysis, with genome synteny and gene loss/gain visually observed. Transcript coverage of select genes was visualised using the Artemis genome browser (v18.0.2) and whole aphid transcriptomic reads generated from ONT

cDNA sequencing. To confirm the presence of amino acid synthesis genes, predicted functional proteins were annotated using GhostKOALA (Kanehisa *et al.,* 2016) to assign KEGG orthology (KO). KO identifiers were uploaded to KEGG to reconstruct the *Buchnera* pathway for biosynthesis of amino acids (map: buc01230).

4.2.21 Sequence alignment of individual Buchnera genes

Orthologous genes and their nucleotide sequence were identified in the Prokka output. Alignments were generated using Mafft (v7.407) (Nakamura *et al.,* 2018), followed by alignment viewing and percent identity calculated using Jalview (Waterhouse *et al.,* 2009).

4.2.22 Phylogenetic analysis of M. euphorbiae and B. aphidicola in the aphid clonal lines

For aphid phylogeny, single copy orthologues were identified between aphid protein sets available from AphidBase, *D. melanogaster* and predicted proteins for *M. euphorbiae* using Orthofinder (v2.2.7) (Emms & Kelly, 2018). For *Buchnera,* single copy orthologues were identified between Prokka annotations for 16 different *Buchnera* strains and *E. coli* (see section 4.2). Orthologous groups were aligned with Mafft (v7.407) (Nakamura *et al.,* 2018) and poorly aligned sequences trimmed using Gblocks (v0.91b) (Castresana, 2002). Concatenated alignments were provided to Prottest (v3.4.2) (Darriba *et al.,* 2011) for model selection. Phylogenetic inference was performed using Raxml-ng (v.0.6.0) (Kozlov *et al.,* 2019) with 1000 bootstraps (see https://github.com/hlmwhite/ProTree_pipe.sh). Phylogenetic trees were output in Newick format and visualised using FigTree (v1.4.3) (Rambaut, 2009).

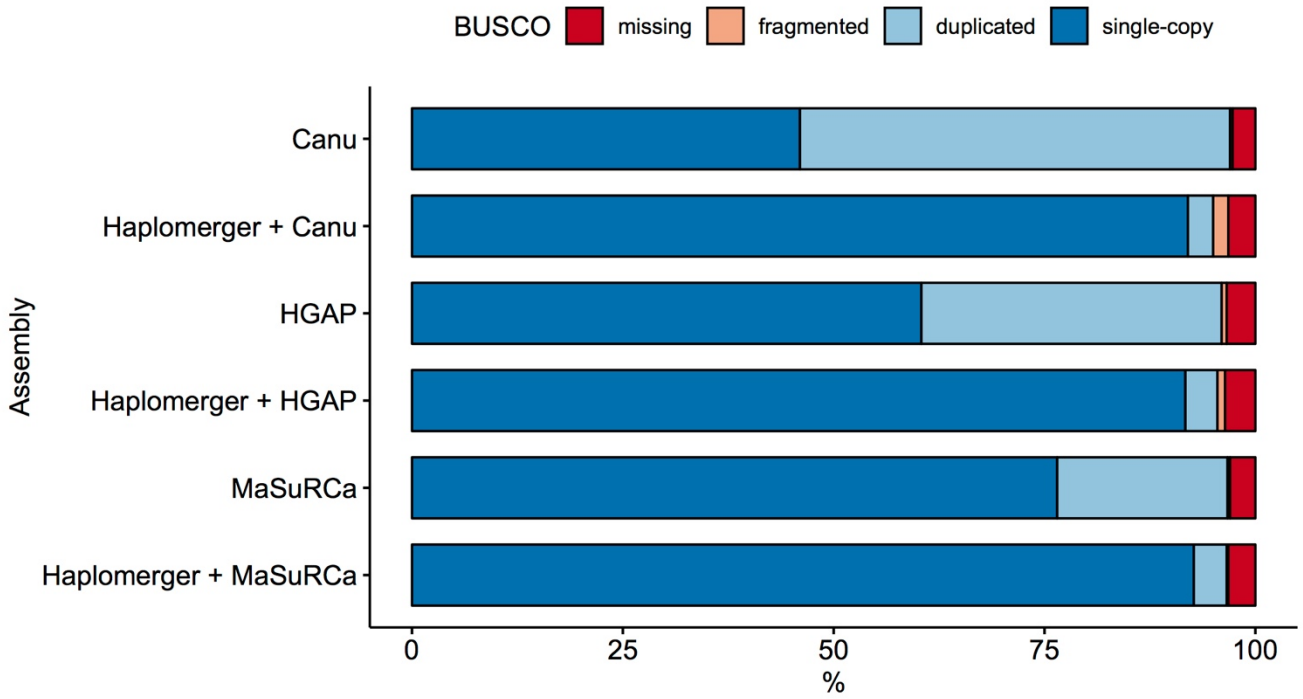
4.3 Results and Discussion

4.3.1 Contig assembly and scaffolding of the M. euphorbiae genome

PacBio sequencing yielded 63.1 Gb of data, equating to 119x coverage of the aphid genome. The three long-read assemblers provided assemblies larger than the predicted 530 mb in Wenger *et al.* (2017) and 526 Mb from GenomeScope analysis (sup. figure 1). GenomeScope also suggests heterozygosity between haplotypes as 1.7%. The Canu assembly was 971.9 Mb, HGAP was 747.4 Mb and MaSuRCa 720.3 Mb. Genome assemblies greater than the predicted size are indicative of multiple haplotypes present within an assembly, which was confirmed through BUSCO analysis and the presence of duplicated orthologues (figure 1). Two rounds of Haplomerger2 were used to reduce the number of haplotypes present in Canu and MaSuRCa assemblies, resulting in a genome size of 560.8 Mb and 550.8 Mb respectively, while only a single round of Haplomerger2 was used to reduce the HGAP assembly to 531.0 Mb . Removing these haplotypes also leads to an increase in contiguity with all three assemblers (figure 2). The Canu assembly N50 increases to 776,046 bp after HaploMerger2, while having the smallest number of contigs out of the three assemblers used (3,159). As a result, the Canu assembly was used for further scaffolding with 10x and Hi-C linkage information.

Why assemblers struggle to successfully collapse diploid assemblies may be rooted in heterozygosity between haplotypes, as seen here in the potato aphid and GenomeScope prediction. This increased heterozygosity has previously been observed in the aphid *R. padi,* where asexual morphs show a microsatellite heterozygosity excess compared to sexual counterparts (Delmotte *et al.,* 2002). With regards to assembly of the potato aphid, removing multiple haplotypes using Haplomerger2 was a necessary step, with a haploid assembly being a requirement for Hi-C scaffolding with HiRise. It is important to note however, that having diploid assemblies may help uncover novel biology, especially when genetic variance in individuals is due to heterozygous alleles.

Blobtools demonstrated very little contamination of the assembly, with 99.5% of sequence matching to *A. pisum* through BLAST, 0.3% match *Buchnera,* 0.2% have no hit and the remainder having < 0.01% matches to other arthropods (sup. figure 3).



genome size	*N50*	*contig num.*
971.9	128,807	10,096
560.8	776,046	3,159
747.4	225,507	6,712
531.0	638,917	3,301
720.3	234,233	6,321
550.8	598,750	3,843

Figure 1. Assessment of arthropod single copy orthologues present in assemblies of *M. euphorbiae.* BUSCO analysis demonstrates how using Haplomerger2 reduces the percentage of duplicated single copy orthologues within assemblies, and that initial assemblies larger than the predicted genome size of 530 Mb comprise multiple haplotypes.

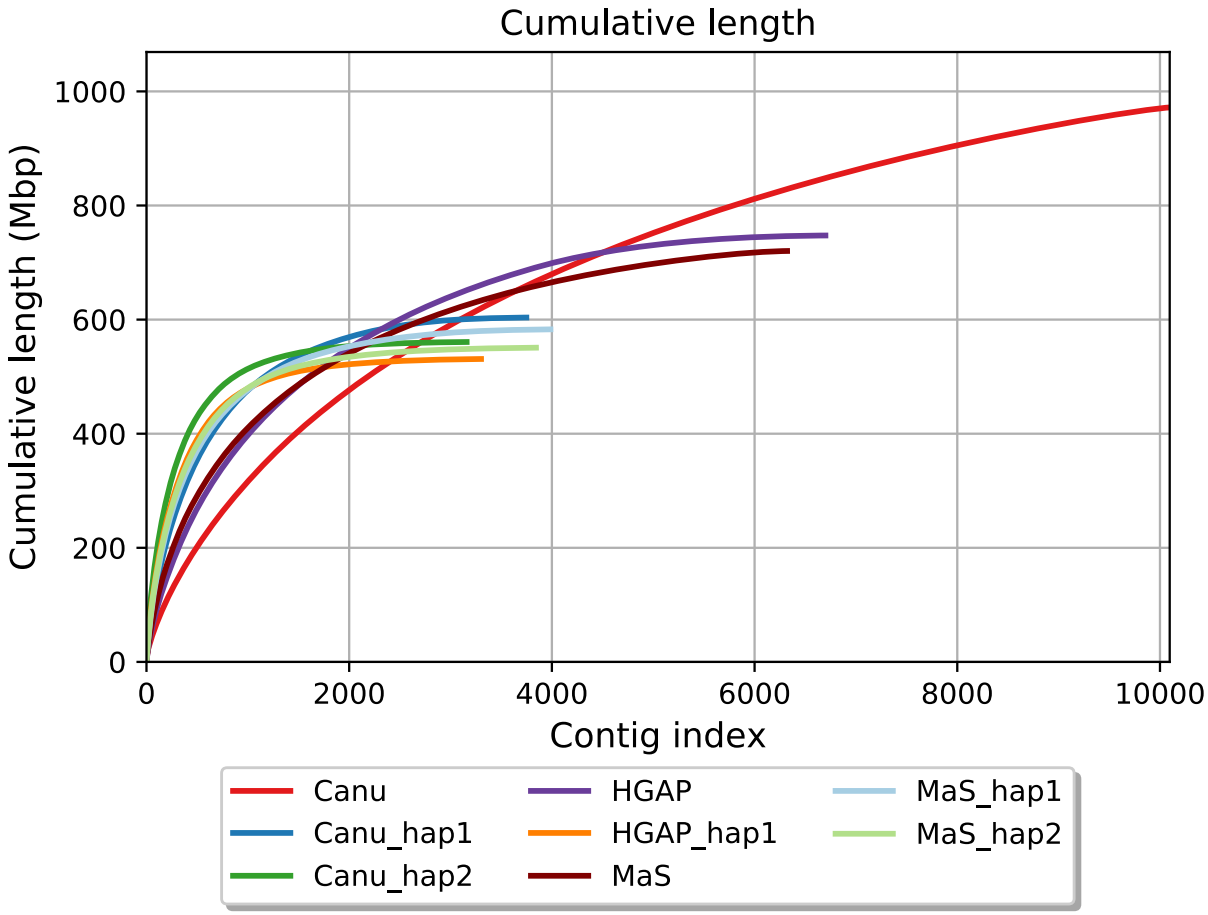


Figure 2. Assembly sizes and their reduction using Haplomerger2. Diploidy is reduced successfully by Haplomerger2, while also showing evidence of increasing assembly contiguity. Appended numbers to assembly names indicate iterations of Haplomerger2 (Graph generated through Quast (v5)).

4.3.2 Genome scaffolding

10x linked reads are useful for phasing and inference of haplotype blocks (Mostovoy *et al.,* 2016), while also useful for further scaffolding the Canu assembly. Increasing the scaffold N50 is also necessary for later orientation and scaffolding with Hi-C. Scaffolding with arcs using 10x reads increases contiguity (3,159 contigs to 2,438 scaffolds), increases the N50 from 776 kb to 2.1 Mb and achieves the desired 1 Mb N50 required for HiRise. Addition of Hi-C linkage information generates a final assembly consisting of 560.8 Mb in 1,918 scaffolds and a scaffold N50 of 57.5 Mb, comparable to other Hi-C assembled aphids (table 1). BUSCO demonstrates the final scaffolded assembly contains 92.5% single copy orthologues, with 2.3% duplicated orthologues and 2.1% fragmented. Of the 1066 orthologues assessed, 3.1% were also missing from the assembly. The largest 14 scaffolds are greater than 1 Mb, with the larger scaffolds reaching chromosome length (longest scaffold = 179.6 Mb) (figure 3). The distribution of BUSCO genes across scaffolds also suggests most gene content is incorporated into these longer sequences (figure 4).

Conserved BUSCO genes between chromosome scale assemblies of *A. pisum* (Li *et al.,* 2019) suggest high levels of gene synteny between the two organisms. *M. euphorbiae* chromosomes 2 (Me 2) and 3 (Me 3) show similarity to *A. pisum* chromosome 1 (Ap 1), sharing 92 BUSCO orthologues and is evidence of a chromosomal split (Figure 5A). While Monti *et al.* (2015) suggest a karyotype n = 5 for *M. euphorbiae*, Hi-C contacts suggest a further chromosome split (figure 3), with Me 1 and Me 5 comprising Ap 2 (274 shared compete BUSCOs) (figure 5A). Other *Macrosiphum* species have the karyotype of 2n = 10 at a minimum (Gavrilov-Zimin *et al.,* 2015), suggesting the chromosomal split is one of the main structural rearrangement events that led to generation of the *Macrosiphum* genus. Gene synteny is not as well conserved between *M. euphorbiae* and *R. maidis* chromosomes (Chen *et al.,* 2019), with much genomic rearrangement (figure 5B). However, Me 2/Me 3 and chromosome Rm 3 are conserved, sharing 93 BUSCO orthologues. As Me 2/Me 3 is also similar to Ap 1, this could suggest

conserved gene synteny across other aphid species. However, chromosome-scale assemblies of other aphid species would be needed to confirm this. Whole genome alignment between *M. euphorbiae* and *A. pisum* further highlights genomic similarity between the two aphid species (sup. figure 2). While chromosomal blocks cluster into similar linkage-groups between each aphid sample, inconsistencies in sequence could be due to genomic rearrangements or misassemblies. Less sequence synteny is observed between *M. euphorbiae* and *R. maidis* (sup. figure 3), where the evolutionary distance between these two species and sequence dissimilarity could be impacting observed nucleotide synteny.

Hi-C greatly improves contiguity, more so than linked-reads alone, where scaffold N50 increases 27-fold, due to Hi-C being able to span longer distances across a chromosome. While 469 Mb of sequence is located within the largest 10 scaffolds, the presence of smaller fragments that remain prior to scaffolding are difficult to place anywhere in the assembly. This equates to 91 Mb of the genome in 1908 scaffolds, all of which are less than 1 Mb. Why these regions are hard to place is unclear, but could be due to remnants of haplotypes not correctly removed during HaploMerger2, a high repetitive content, or low coverage of mapped 10x or Hi-C reads that therefore does not provide enough linkage information to correctly scaffold. Being able to incorporate as much sequence within assembled chromosomes may be crucial in downstream analysis, especially if an observed phenotype is the result of large structural variation such as chromosomal inversions (Kirkpatrick & Barton, 2006).

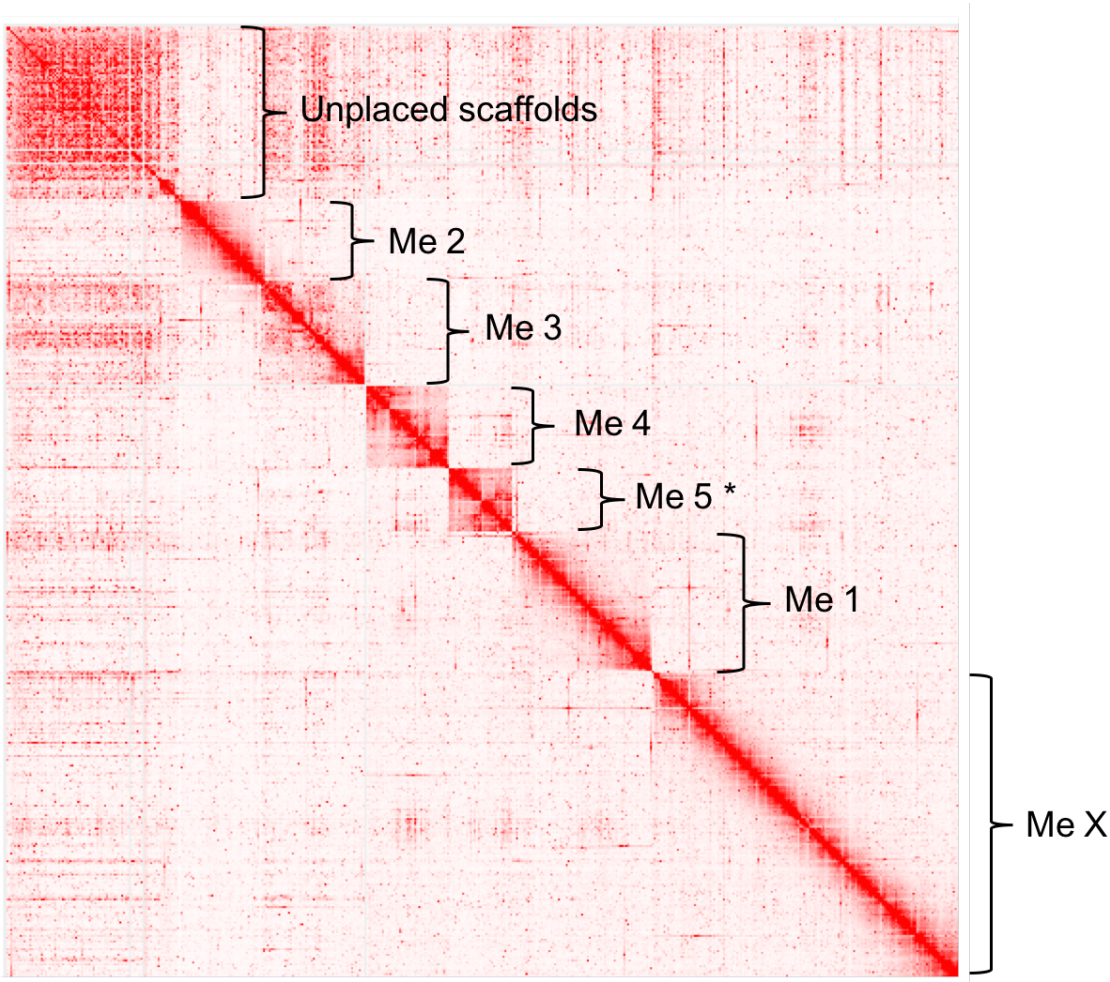


Figure 3. Hi-C contact map of the HiRise assembly. Large scaffolds indicate some full-length or near full-length chromosomes. However, many small unplaced contigs at the top are difficult to orient and place correctly within the assembly. Hi-C contacts also suggest a further chromosome split with Me 5*. Graph generated using Juicebox.

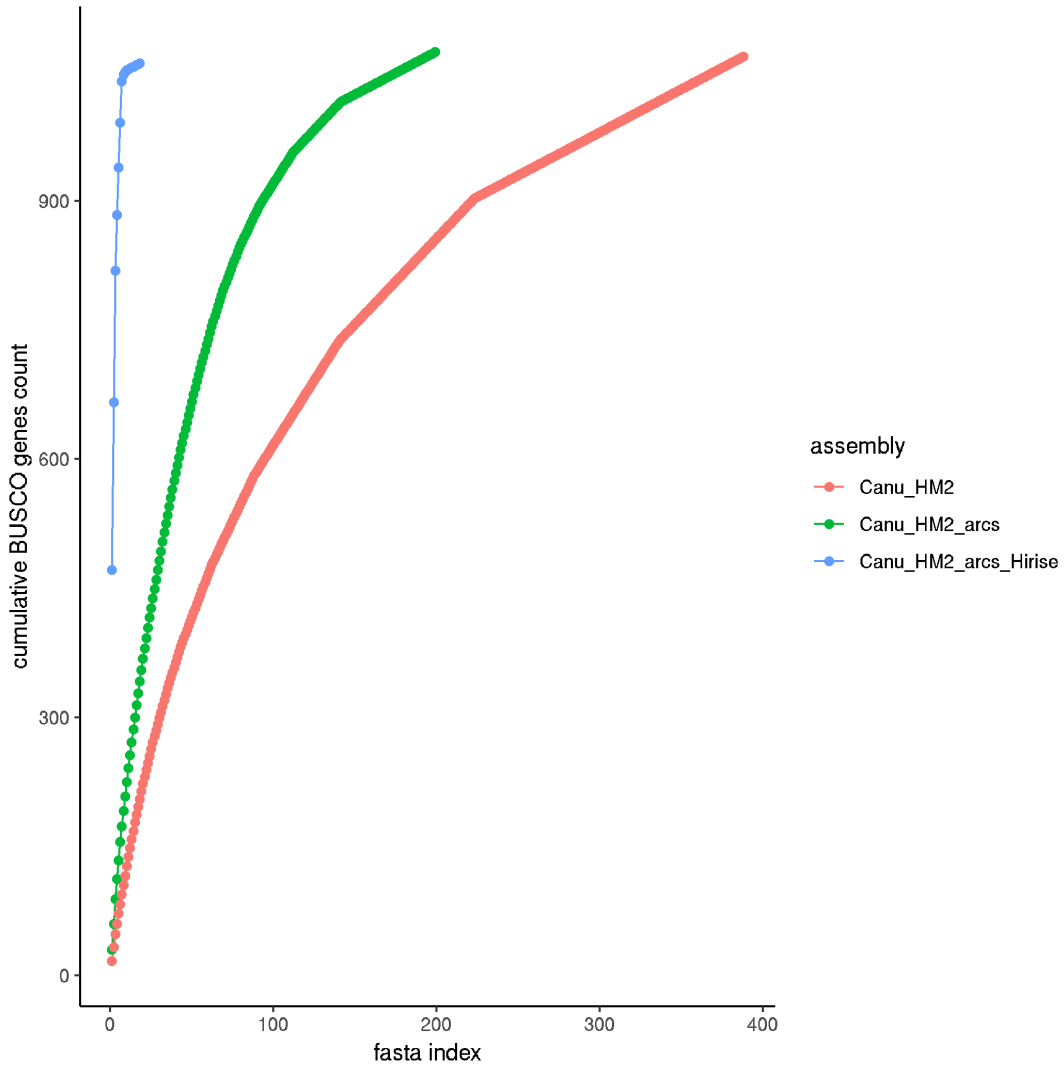
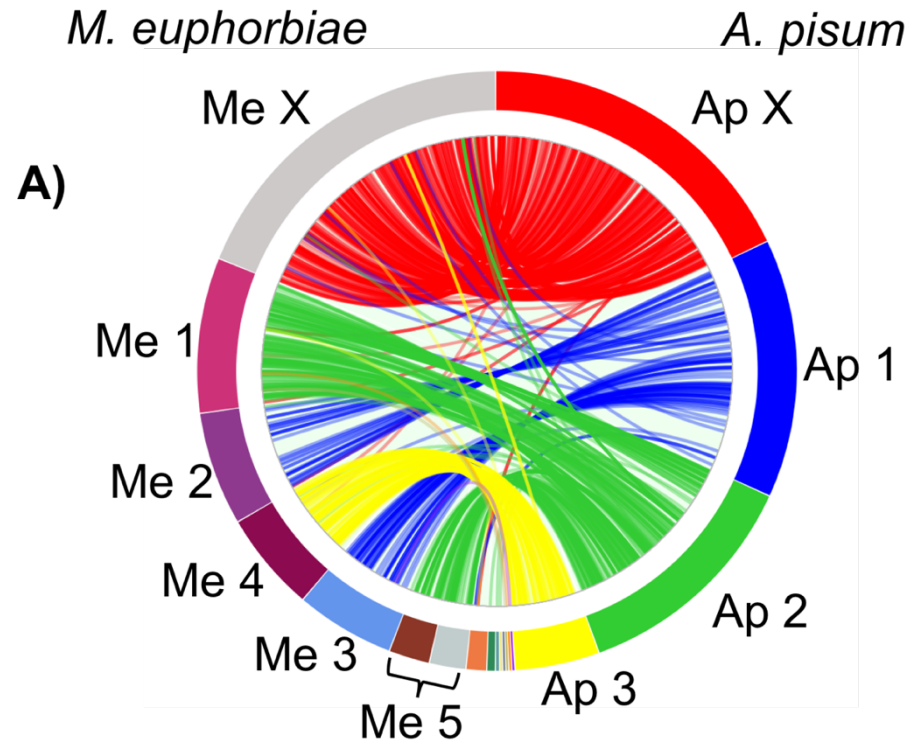


Figure 4. BUSCO score content related to increasing scaffold size demonstrates the incorporation of genes into fewer scaffolds. Scaffolds were ordered in size from largest to smallest and plotted against the number of complete single arthropod BUSCOs per fasta entry. The three assemblies shown here are the initial haploid assembly generated from Canu and Haplomerger2 (Canu_HM2), followed by scaffolding first with 10x linked reads and arcs (Canu_HM2_arcs) and followed by further scaffolding with Hi-C reads (Canu_HM2_arcs_Hirise).


M. euphorbiae
A. pisum
A)
Me X
Ap X
Me 1
Ap 1
Me 2
Me 4
Ap 2
Me 3
Ap 3
Me 5

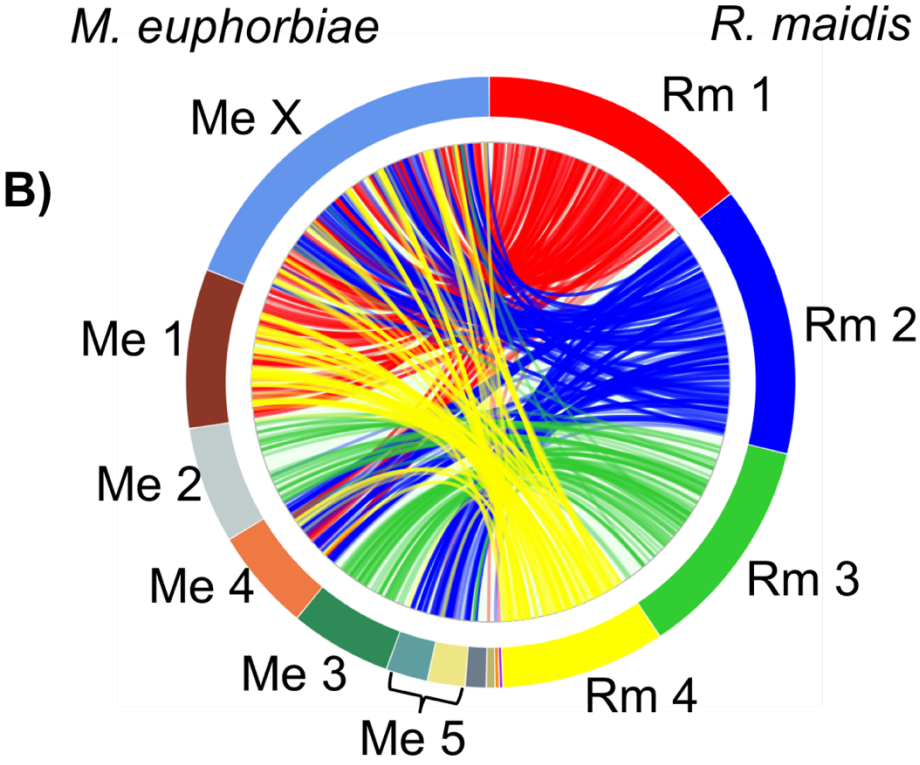

M. euphorbiae
R. maidis
B)
Me X
Rm 1
Me 1
Rm 2
Me 2
Me 4
Rm 3
Me 3
Rm 4
Me 5

Figure 5. Shared complete arthropod BUSCO genes between chromosomes of A) *A. pisum* and *M. euphorbiae* (BUSCO n = 965), and B) *R. maidis* and *M. euphorbiae* (BUSCO n = 963). Each genome is represented by either the left or right hemisphere of each plot. Scaffolds with complete BUSCO genes are shown here, therefore 465.2 Mb, 467.2 Mb and 321.0 Mb of *A. pisum, M. euphorbiae* and *R. maidis* are represented here. 'X' represents X chromosomes for *M. euphorbiae* and *A. pisum*.

4.3.3 Repetitive elements of M. euphorbiae

RepeatModeler and RepeatMasker reported 38.06% of the genome consists of repetitive elements in *M. euphorbiae*. This also conforms with GenomeScope analysis, which predicts 44.4% of the genome to consist of repetitive elements (sup. figure 2). Repeat analysis of *A. pisum*, *M. persicae* (G006 clone) and *R. maidis* contain fewer repeat sequences compared to *M. euphorbiae* (table 2). An LTR-*gypsy*-like retrotransposon is at least four times as common in *M. euphorbiae* (1.48%) compared to *A. pisum* (0.29%) (sup. Table 1), with LTR-*gypsy* more commonly observed in plant species (Schnable *et al.,* 2009; Qiu & Ungerer, 2018; Wicker *et al.,* 2018). RC helitron and DNA transposon Mavericks are also more prevalent compared to other aphid species, the former of which implicated in shaping genome architecture and horizontal gene transfer in insects (Han *et al.,* 2019) (see http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/repeats_table_aphids_2019.csv). Approximately 40 mb of the *A. pisum* assembly are gapped Ns, and could be masking repetitive regions from being discovered through RepeatMasker, and hindering comparative analysis. Within identified repeat regions, previously described telomeric repeats ((TTAGG)n) were not apparent at any chromosome end (Monti *et al.,* 2011), possibly due to removal of highly repetitive sequences during the Canu assembly step.

RepeatMasker identifies many repetitive regions within *M. euphorbiae*, potentially taking up 40% of the whole genome. The high proportion of repeats highlights the necessity of using long reads for assembly, where short reads alone

are unable to span long repetitive regions. The ability to handle repeat regions correctly can also be useful in the discovery of more gene content, such as the identification of more transposable elements, or identifying complex variants such as gene duplications. The latter especially important regarding other Macrosiphini species, such as *A. pisum* and *M. persicae*, which show extensive gene family expansion (IAGC, 2010; Mathers *et al.,* 2017).

For *M. euphorbiae*, transposable elements (TEs) are more frequent compared to *R. maidis* and *M. persicae*, while being more comparable to proportions observed in *A. pisum* (IAGC, 2010). This could be an effect of long-read sequencing increasing TE discovery, where they are able to capture TEs nested within long repetitive regions (Disdero & Filée, 2017). TE accumulation in asexual species may occur, where unchecked proliferation of TEs are not selected out through sexual recombination (Dolgin & Charlesworth, 2006), and could be an indicator of obligate parthenogenesis in UK potato aphid populations.

Table 2. Repeat element comparison of four aphid genomes following RepeatModeler/RepeatMasker. *M. euphorbiae* and *A. pisum* show higher amounts of repetitive content compared to *M. persicae* and *R. maidis.* LINEs and SINEs are long interspersed nuclear elements and interspersed nuclear elements respectively. LTRs are long terminal repeats. '*' *Repeat proportions differ to those reported in the Pea aphid genome paper (IAGC, 2010) where it was calculated using a reduced assembly size of 464 Mb and 38% repetitive sequence.*

	M. euphorbiae	***A. pisum****	***M. persicae* (G006)**	***R. maidis***
bases masked %	38.06	29.55	18.50	19.53
SINEs %	0.00	0.00	0.00	0.00
LINEs %	1.96	1.31	1.04	0.54
LTR elements %	2.16	0.53	0.33	0.09
DNA elements %	3.36	1.51	1.14	1.04
Unclassified %	27.20	22.85	12.11	13.02
Total interspersed repeats %	34.69	26.19	14.63	14.69
Small RNA %	0.00	0.00	0.00	0.00
Satellites %	0.00	0.00	0.00	0.00
Simple repeats %	3.10	3.06	3.45	4.25
Low complexity %	0.36	0.36	0.44	0.61

4.3.4 Transcriptome assembly and gene prediction with MinION and Illumina transcriptomic data

MinION sequencing provided 1,227,246 reads, equating to 1.5 Gbp. The low yield of MinION reads results in a small number of assembled transcripts, especially when the minimum cluster size for pinfish is set higher. While setting a low cluster size of two increases the number transcripts and full length transcripts retained, 29,370 are still considered fragmented and incomplete when compared to the most related BLAST hit in a swissprot database. The proportion of complete to fragmented transcripts increases with a higher cluster threshold (table 3), although this also reduces the number of full length genes.

Transcriptomic reads from Teixeria *et al.* (2018) (SRA accession SRX339176) consists of 274,454,253 reads and 9.2 Gbp, which was used for transcriptome

assembly with Trinity. The Trinity assembly provides a higher number of assembled transcripts and is more complete based on BUSCO scores (86.6% complete). This is in comparison to MinION and pinfish assembled transcripts, where complete BUSCO scores for cluster thresholds of 2, 5 and 10 are 34.5%, 11.65% and 6.2% respectively.

Table 3. MinION and Illumina assembled transcripts. Cluster threshold defines the smallest number of mapped transcripts required to be classed as a gene. Full length transcripts are assessed through BLAST against a swissprot database and binned based on match length using 'analyze_BLASTPlus_topHit_coverage.pl' (Trinity package).

	cluster threshold	*assembled transcripts*	*full length transcripts*
MinION (pinfish)	2	29,842	473
	5	7,268	213
	10	3,312	151
Illumina (Trinity)	N/A	69,957	1,998

Transcriptome assembly was necessary for genome annotation with Maker2, where either all assembled transcripts were provided, or just full-length transcripts. Following filters for pfam domains and AED scores below 1, either method for annotation with Maker2 yields an insufficient number of genes based on BUSCO scores (figure 6). Gene annotations predicted using Maker2 were either larger than estimated (51,429 – all transcripts) or were missing core orthologues (88.6% complete BUSCOs – full-length transcripts only). Having an inflated gene number indicates a high rate of false positives and may affect further analysis such as studying copy-number variation, while low BUSCO scores indicate incomplete gene content and may lack genes of interest.

BRAKER was provided with mapped transcriptomic reads, bypassing the requirement for transcriptome assembly. Compared to Maker2, BRAKER provided improved gene calls, with any combination of input data giving consistent results between them based on BUSCOs (all greater than 93.8% complete) and gene number (figure 6). However, number of predicted genes is high (42,952). Genes were then selected on the presence of mapped transcriptomic evidence, which was necessary to remove falsely called genes as this may impact analysis of gene copy number. Following selection, gene count falls to 26,802 and agrees with gene predictions in other aphid species (Thorpe *et al.,* 2018). BUSCO indicates this final gene set is still near complete, having 94.5% complete arthropod orthologues while missing only 2.7% (figure 6). InterProScan annotated 15,169 predicted gene-encoded proteins with functional pfam domains. Not every protein may be annotated due to a lack of specific aphid hmm references not present in the Pfam database.

The BRAKER pipeline utilizing published Illumina RNA-seq reads provided better gene calls for further study and is supported by near-complete gene sets as assessed through BUSCO (94.6% complete). Surprisingly, the MinION evidence alone when provided to BRAKER generated similar results in terms of gene numbers and BUSCO scores (42,900; 93.8% complete). Read-mapping through the splice aware Minimap2 may provide better hints for training AUGUSTUS rather than mapping assembled transcripts to the genome as seen in Maker2. In the case of MinION transcripts, these are more likely to be error-prone as some genes will have inadequate coverage for transcript polishing during pinfish assembly. Erroneous reads may be more difficult to map through BLAST and exonerate during Maker2 annotation and therefore insufficient for gene calling.

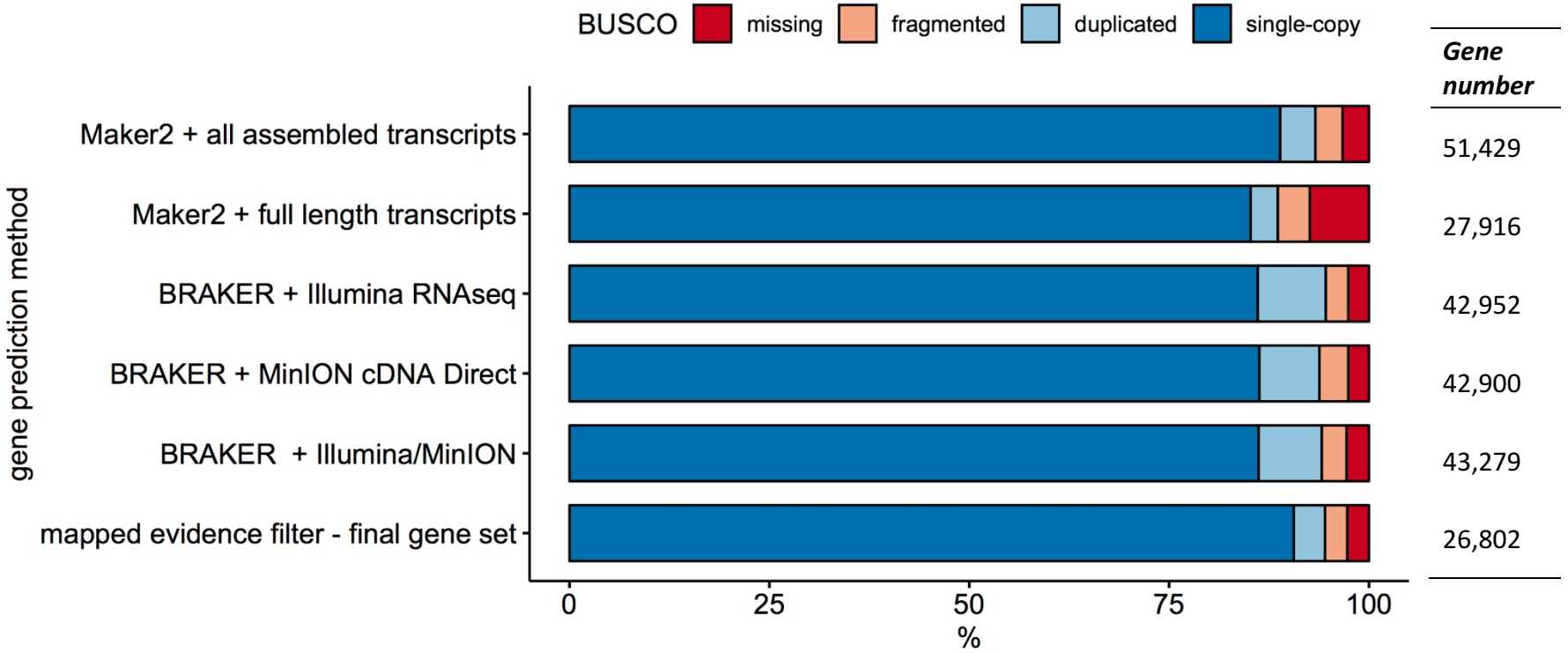


Figure 6. Gene number predictions using Maker2 and BRAKER with transcriptome completeness assessed using arthropod BUSCOs. Filtered gene set consists of predictions from the BRAKER with mapped Illumina transcriptomic evidence observed through featureCounts.

4.3.5 Phylogenetic analysis of M. euphorbiae

Between aphid species and *D. melanogaster*, Orthofinder identified 387 single copy orthologues to use in generating phylogeny. Tree inference with raxml-ng used a JTT+G+F model as predicted with prottest3. *A. pisum* is the most closely related species of *M. euphorbiae* currently sequenced, which agrees with previously described genome synteny and genome size observed between both aphid species (IAGC, 2010) (figure 7).

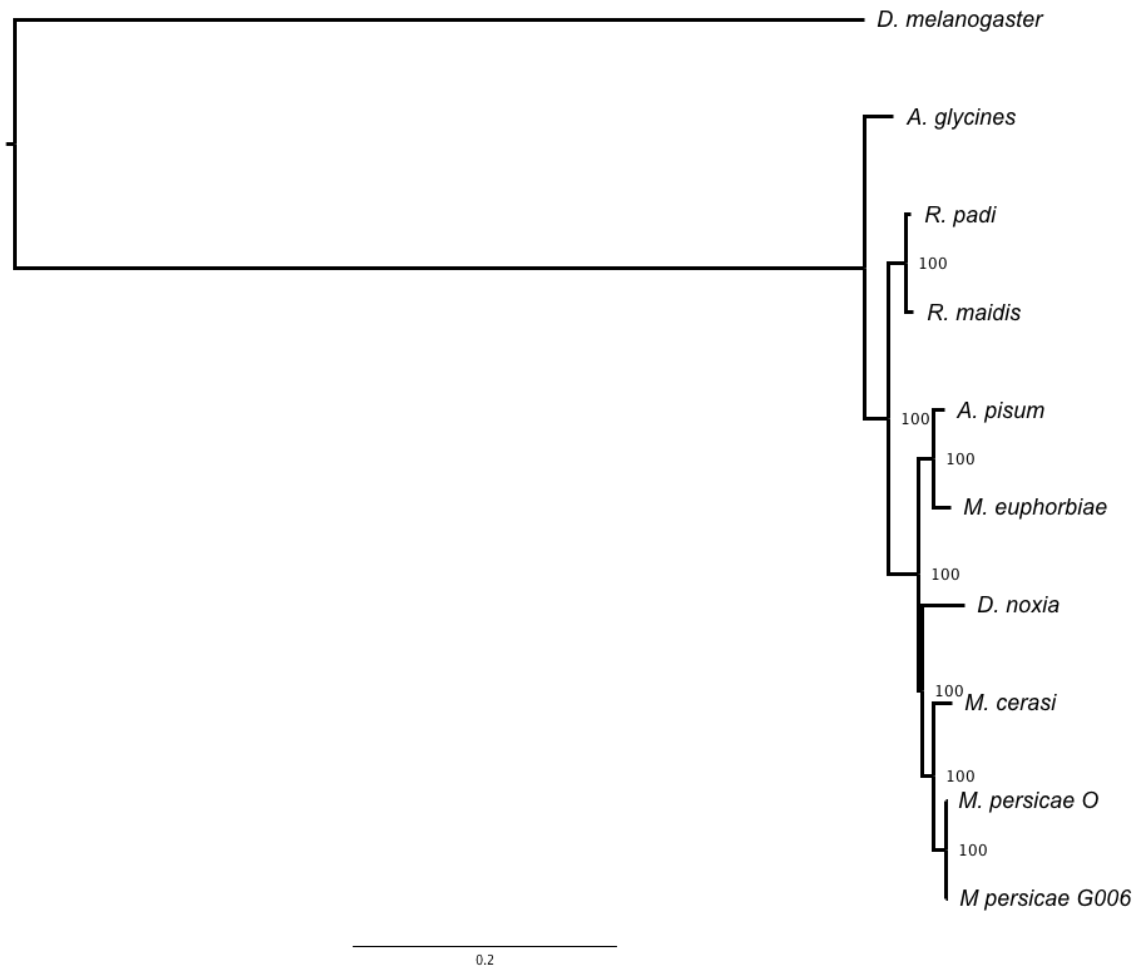


Figure 7. Phylogenetic analysis of aphid species. Tree was created using a maximum-likelihood method with 1000 bootstraps. *D. melanogaster* was used as an outgroup. Node numbers indicate support values. Scale bar indicates amino acid substitutions per site.

4.3.6 Shared immune function proteins between M. euphorbiae and A. pisum

Through Orthofinder analysis, 111 potato aphid proteins shared orthology to immune related proteins outlined in Gerardo *et al.* (2010) and Ma *et al.* (2019) (table 4). While gene numbers are similar between most groups, some immune function groups stand out for *M. euphorbiae*. *M. euphorbiae* possess more genes related to 'general stress responses' (n = 20) than any other aphid genome assessed here, and encode heat shock proteins (Hsps) or heat shock cognates (Hscs). HSPs are proteins conserved across all organisms that have a wide range of function, from protein folding, cell signalling, cell defence against stress as well as roles in immune activation (Li & Srivastava, 2003; Pockley, 2003). Specifically, the potato aphid encodes nine orthologues of heat shock cognate 70-1 (Hsc70-1), more than any other aphid studied here (*A. pisum, M. persicae, D. noxia* and *A. glycines* encode 5, 4, 8 and 3 respectively). (More detailed output of aphid immune orthologues are deposited at http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/aphid_immune_orthologues.csv).

Hsc70-1 and heat shock protein 70 (Hsp70) have been extensively linked to stress and cold tolerance (Burton *et al.,* 1988; Reinhart *et al.,* 2000; Reinhart *et al.,* 2007; Koštál, V & Tollarová-Borovanská, 2009; Sun *et al.,* 2016) as well potentially in pesticide resistance (Sun *et al.,* 2016). Both these factors are especially noteworthy for *M. euphorbiae,* where cold tolerance in genotype 1 is hinted at in the previous chapter and other work into genotype 1 insecticide resistance (Clarke *et al.,* 2018). Regarding pesticide resistance, another important family of immune genes is those linked to detoxification, consisting of genes encoding Glutathione-S-transferases (GSTs). GSTs are responsible for the detoxification of xenobiotic compounds via conjugation to glutathione, resulting in compounds much more soluble and easier to excrete from the cell. GST also helps remove harmful reactive oxygen species (ROS) in this manner that accumulate during times of cell stress (Veal *et al.,* 2002; Enayati *et al.,* 2005). However, whether Hsps and GST play a role directly in parasitoid resistance, or simply allow the aphid to tolerate

the stress of *A. ervi* challenge and improves efficacy of other immune pathways remains to be seen. One such pathway may be phenoloxidase pathways, which results in melanisation of foreign bodies introduced to the aphid (Schmitz *et al.,* 2012). Potato aphids potentially have more clip-domain serine protease orthologues (table 4), with 3 orthologues of *Ap_VP* and 5 orthologues of *AP_SPLP*. Whether increased gene copy number plays a role in modulating expression of melanisation and phenoloxidase response remains to be seen, but could implicate it in parastioid resistance if variation is further observed in other genotypes of the Potato aphid (see chapter 5).

Table 4. Predicted functional groups involved in aphid immune defence, as taken from Gerardo *et al.* (2010) and Ma *et al.* (2019). Counts indicate number of genes identified within each immune function group for four aphid species using a list of encoded Pea aphid predicted immune proteins.

predicted immune function	*A. pisum*	*M. euphorbiae*	*M. persicae (O clone)*	*D. noxia*	*A. glycines*
alarm pheromone production	2	2	2	2	3
antimicrobial peptide	7	5	4	2	1
bacterial and fungal pattern recognition	2	2	2	4	2
bacterial recognition, induction of phenoloxidase	5	4	3	4	4
detoxification	19	15	13	9	11
fungal degradation	8	7	8	10	6
general stress response	16	20	15	17	12
IMD pathway	4	5	4	4	4
jak/stat pathway	9	5	7	4	6
Janus kinase pathway	7	6	5	3	4
mark pathogens for phagocytosis	2	1	1	1	1
microbial degradation	3	3	3	3	3
multiple hypothesized functions	2	2	2	2	1
produce nitric oxide, a toxic gas	1	1	1	1	1
proPO 1/2	2	1	2	2	2
clip-domain serine protease	4	8	3	3	3
toll pathway	26	24	17	20	18

4.3.7 Genome assemblies for B. aphidicola of M. euphorbiae

The genome of *B. aphidicola* is highly reduced due to its evolved mutualism with the host aphid (Van Ham *et al.,* 2003; Wilson *et al.*, 2010). Blobtools analysis indicate the short read DISCOVAR assembly for AA09/04 contains a complete *B. aphidicola* sequence (sup. figure 3). *B. aphidicola* consists of a single circular chromosome (645,270 bp), along with its two plasmids pLeu (7,754 bp) and pTrp (3,529 bp) (figure 8). Prokka predicts 625 genes (table 5), with pseudofinder analysis indicating 565 Prokka predicted functional genes are present on the *Buchnera* chromosome, similar to predicted gene numbers for *Buchnera* symbionts of *A. pisum* (n = 567) and *M. euphorbiae* sequence available from NCBI (n = 566). *B. aphidicola* was found to have all genes relevant to amino acid biosynthesis previously identified in *Buchnera* from *A. pisum* through annotation with KEGG orthology annotation (sup. figure 5).

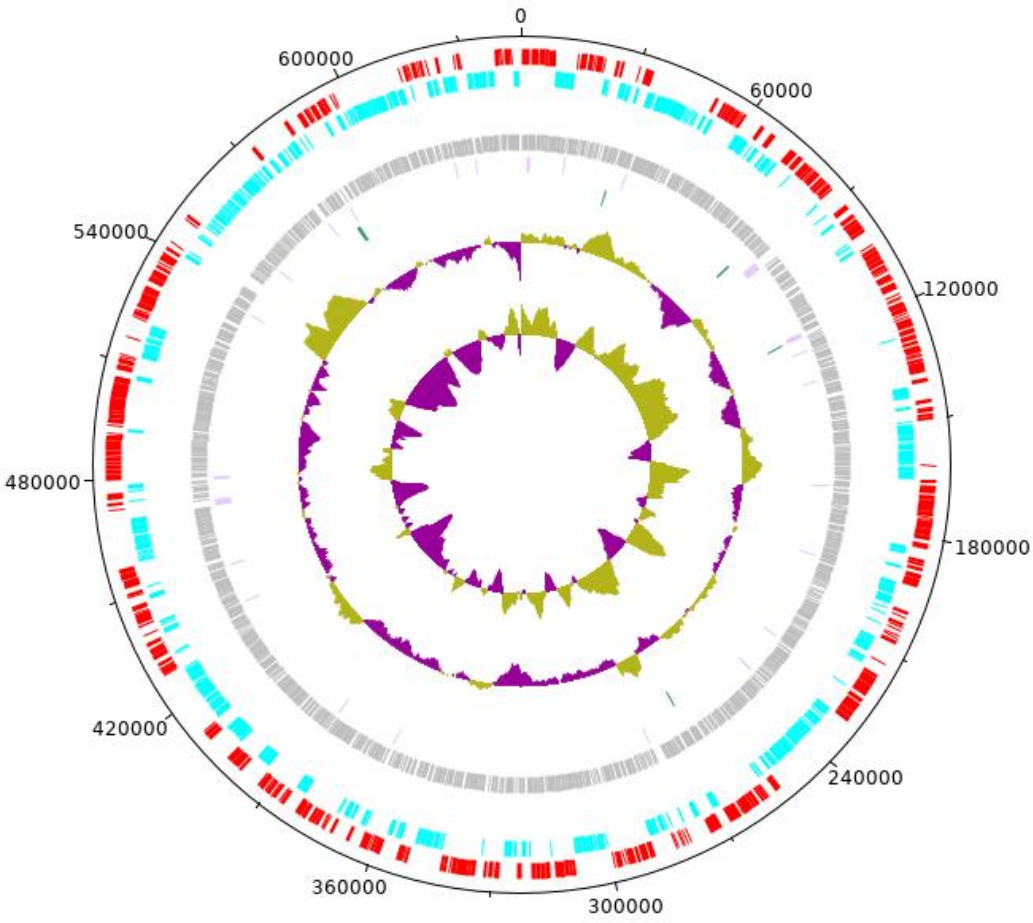
Analysis of *Buchnera* strains through ACT belonging to *A. pisum* and *M. euphorbiae* indicated strong levels of conserved synteny overall, which is also reflected through phylogenetic analysis (figure 9). However, the *Buchnera* of *A. pisum* appears to be lacking four genes observed in the *M. euphorbiae* counterpart: *metR, fliK, fabZ* and *ansA.* Also of interest is reduced sequence similarity in the riboflavin synthase gene *ribC* compared with the *Buchnera of A. pisum* (figure 9 and sup. figure 6), where only 84.06% identity is observed. In comparison, there is 98.72% identity shared between *ribC* belonging to *B. aphidicola* from both genotypes of *M. euphorbiae*. Riboflavin is predicted to have an impact on aphid nymph development, especially in embryonic stages where it is upregulated (Nakabachi, & Ishikawa, 1999; Bermingham *et al.,* 2009). Genes involved in riboflavin biosynthesis and transport could be a source of potential phenotypic variation for the bacterial symbiont and its host. Variation between *rib* genes allows discrimination between species of the *Bartonella* bacterium (Bereswill *et al.,* 1999), and inactivation of *ribC* results in riboflavin overproduction (Vitreschak *et al.,* 2002) (riboflavin provision and inter-genotype variation in Potato aphids is further discussed in chapter 5). *metR* and *ansA* are involved in

methionine regulation and asparagine metabolism respectively. *metR* and *ansA* are believed to be either non-functional or not present within *the Buchnera* of *A. pisum* (Moran *et al.,* 2005; Jiang *et al.,* 2013), but are within the *Buchnera* of *M. persicae* (Jiang *et al.,* 2013). A third strain belonging to *Acyrthosiphon kondoi* is one of the only other *Buchnera* genomes to encode *ansA*, further suggesting correct placement of *M. euphorbiae Buchnera* within the phylogeny (figure 10). Transcript coverage of *Buchnera* suggests functional *ansA* (sup. figure 7A) but little coverage suggests *metR* is not, or is very tightly regulated (sup. figure 7B). Jiang *et al.* suggest the presence of genes such as *ansA* and *metR* permit a broader host range through the ability to utilize more of the amino acid content of host plant phloem sap. This could explain why *M. euphorbiae* are polyphagous; further study of the *Buchnera* genotype between potato aphid genotypes could explain disparities in performance between aphid genotypes on different host plant species (Karley *et al.,* 2017).

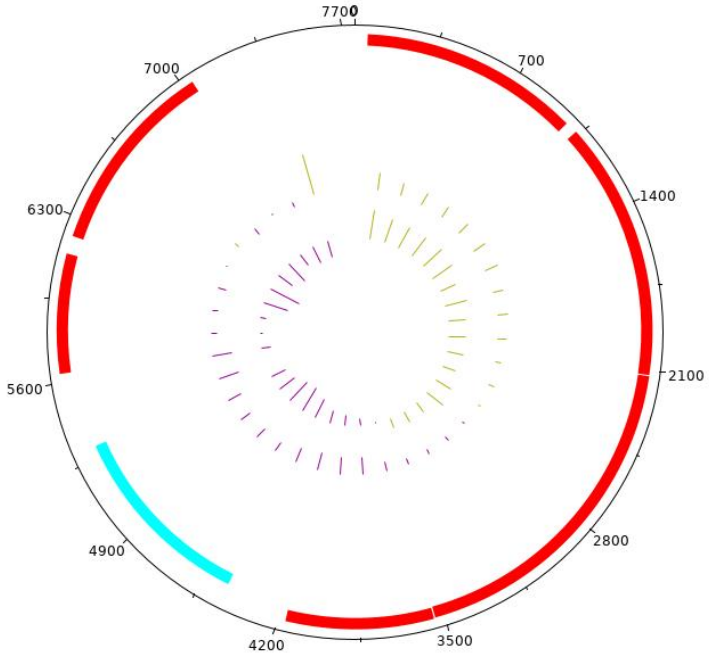
Table 5. Comparison of gene content for three *B. aphidicola* genomes. Total predicted genes were generated using Prokka (Seemann, 2014). Genes which encode functional proteins as well as pseudogene candidates were generated using psuedofinder (Syberg-Olsen & Husnik, 2018). Genes were re-predicted as part of the study to for comparative purposes between genomes. * *taken from NCBI (NCBI assembly GCA_005237295.1).*

Buchnera aphid host	***Prokka predicted total genes***	***Buchnera chromosome functional genes***	***predicted psuedogenes***	***pTrp plasmid genes***	***pLeu plasmid genes***
A. pisum	638	567	19	4	7
*M. euphorbiae**	620	566	25	(not present)	(not present)
M. euphorbiae	625	565	24	4	7

A)


0
60000
120000
180000
240000
300000
360000
420000
480000
540000
600000

B)


7700
700
1400
2100
2800
3500
4200
4900
5600
6300
7000

C)

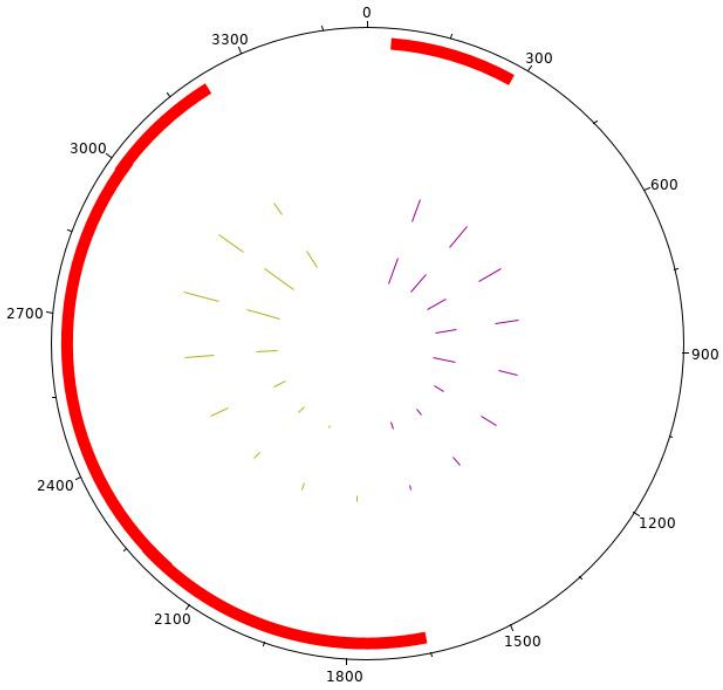

0
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300

Figure 8. A) Genome of *B. aphidicola*; B) pLeu plasmid; C) pTrp plasmid. Outer-most track 1) Forward strand coding sequence (CDS); 2) Reverse strand CDS; 3) Source feature track; 4) Purple dashes represent pseudogenes predicted through pseudofinder; 5) Green dashes represent position of genes predicted in *B. aphidicola* (*M. euphorbiae*) but not *B. aphidicola* (*A. pisum*); 6) GC plot; 7) GC skew.

Figure 9. Nucleotide sequence alignment of *ribC* gene of *B. aphidicola* from *A. pisum, M. euphorbiae* (NCBI download) and *M. euphorbiae* in this study. Lack of sequence identity between *ribC* in *Buchnera* from these two aphid species could indicate a source of phenotypic variation, or at least highlights the gene as being polymorphic between *Buchnera* strains.

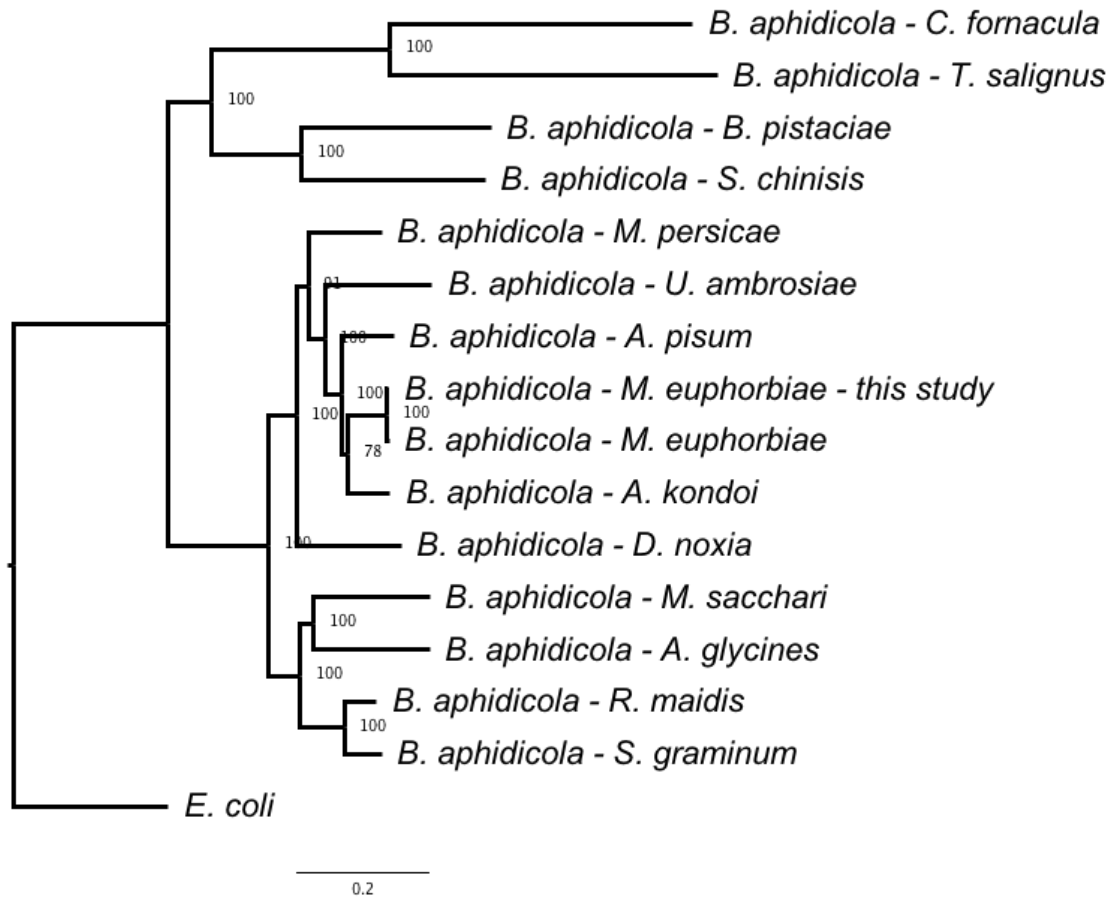


Figure 10. Phylogenetic analysis of current *B. aphidicola* strains. Phylogeny was inferred from 293 single copy orthologues across all 16 genomes. Tree was created using a maximum-likelihood method with 1000 bootstraps, using a CpREV+I+G+F model for amino acid substitution. *E. coli* was used as an outgroup. Node numbers indicate support values. Scale bar indicates amino acid substitutions per site.

4.4 Conclusion

Here is presented a near chromosome scale assembly of the potato aphid *M. euphorbiae*, as well as its obligate symbiont *B. aphidicola*. Features of this genome, such as high repetitive content, demonstrate the value of long read sequencing and long-range information in the form of Hi-C and 10x linked reads. While *Buchnera* assembled into a single contig and its two plasmids, the presence of small unplaced scaffolds remains an issue for complete assembly of the aphid host. Gene synteny is conserved between *M. euphorbiae* and *A. pisum,* but indicates a chromosomal split that generates *Macrosiphum* chromosomes 2, while Hi-C contacts suggest a further chromosomal split and karyotype (n = 6). This would however, require further cytogenetic analysis to confirm.

Gene prediction for both aphid and symbiont was also challenging, especially for the aphid host. MinION sequencing may still not be perfectly suited for eukaryote transcriptomic analysis, but it complemented previously generated Illumina data also analysed here, and provided transcriptome evidence for *Buchnera.*

The following chapter will use predicted genes of interest for studying observed genotype 1 phenotypes, such as parasitoid resistance and aphid response to environmental stress.

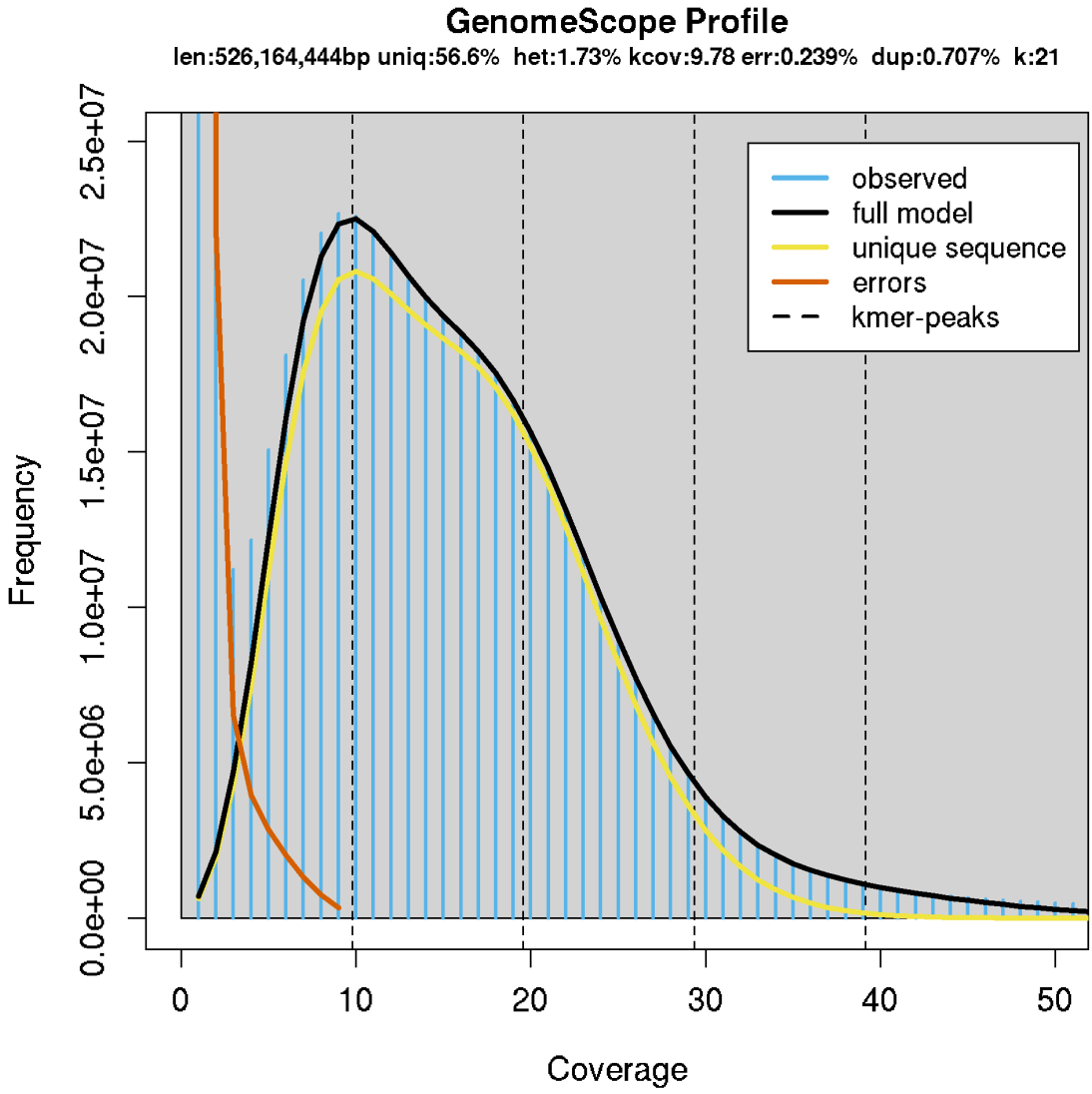
4.5 Supplementary tables

Supplementary table 1. Repeat elements of *M. euphorbiae* as predicted through RepeatModeler and RepeatMasker..

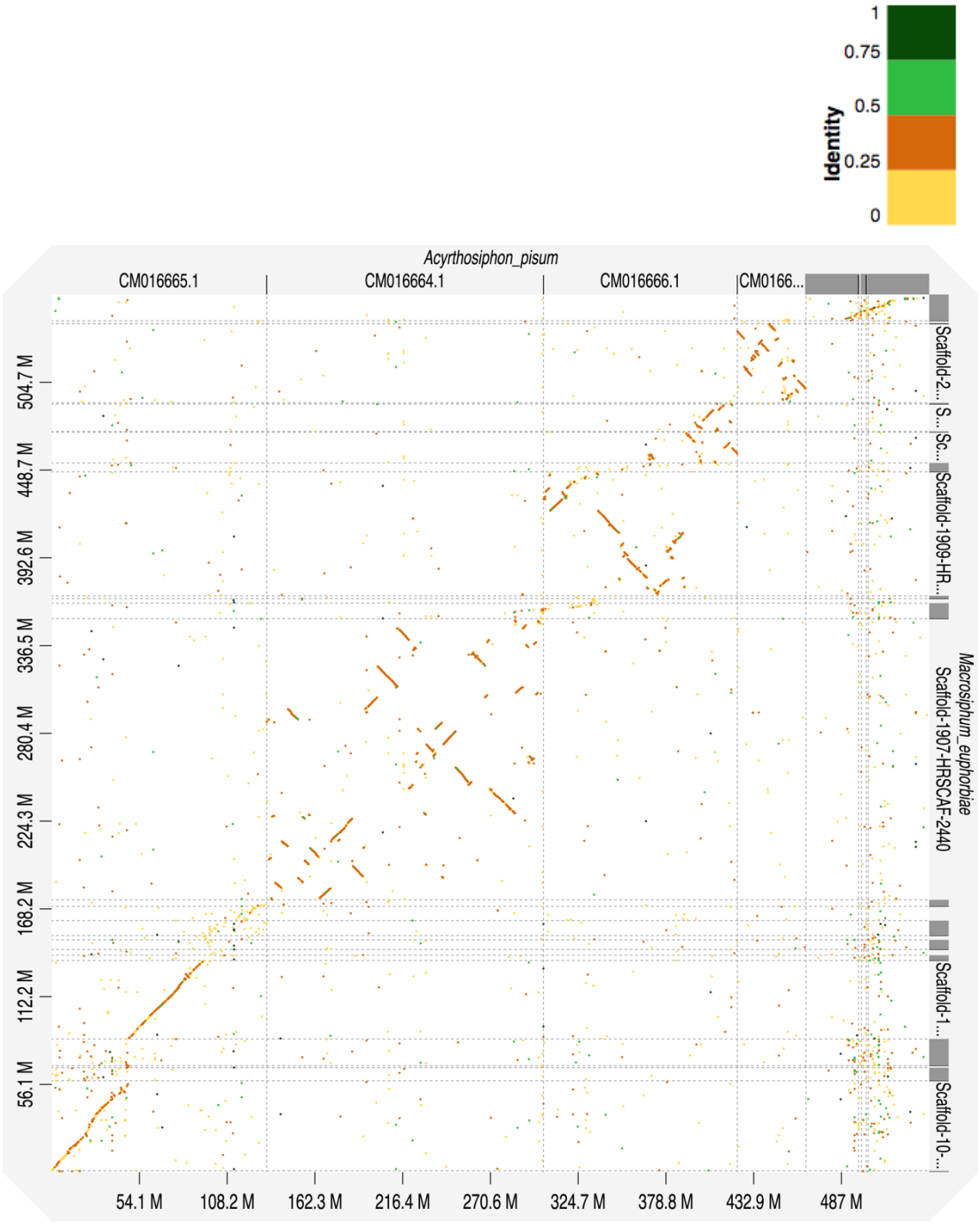
Transposable element	*bp*	*percentage of genome (%)*
DNA	146741	0.03
DNA_Academ	114984	0.02
DNA_CMC-EnSpm	198808	0.04
DNA_Ginger	428342	0.08
DNA_hAT	631821	0.11
DNA_hAT-Ac	775617	0.14
DNA_hAT-Blackjack	61745	0.01
DNA_hAT-Charlie	502292	0.09
DNA_hAT-hATm	573065	0.10
DNA_hAT-Pegasus	43642	0.01
DNA_hAT-Tip100	3374541	0.60
DNA_hAT-Tol2	76517	0.01
DNA_Kolobok-Hydra	307744	0.05
DNA_Maverick	7887036	1.41
DNA_Merlin	259064	0.05
DNA_MuLE-MuDR	1559942	0.28
DNA_MuLE-NOF	332453	0.06
DNA_P	904832	0.16
DNA_PIF-Harbinger	511906	0.09
DNA_PIF-ISL2EU	99610	0.02
DNA_Sola	168380	0.03
DNA_TcMar-m44	358685	0.06
DNA_TcMar-Mariner	53798	0.01
DNA_TcMar-Tc1	156889	0.03
LINE_CR1	1683867	0.30
LINE_I	899808	0.16
LINE_I-Nimb	315587	0.06
LINE_Jockey	3302853	0.59
LINE_L1-Tx1	106155	0.02
LINE_L2	1602375	0.29
LINE_R1	1643166	0.29
LINE_R2	3334	0.00
LINE_RTE-BovB	1855730	0.33
Low_complexity	1973809	0.35
LTR	578945	0.10
LTR_Copia	1302478	0.23
LTR_Gypsy	8326660	1.48

LTR_Pao	2211509	0.39
RC_Helitron	3091313	0.55
Simple_repeat	17200324	3.07
Unknown	160414897	28.60

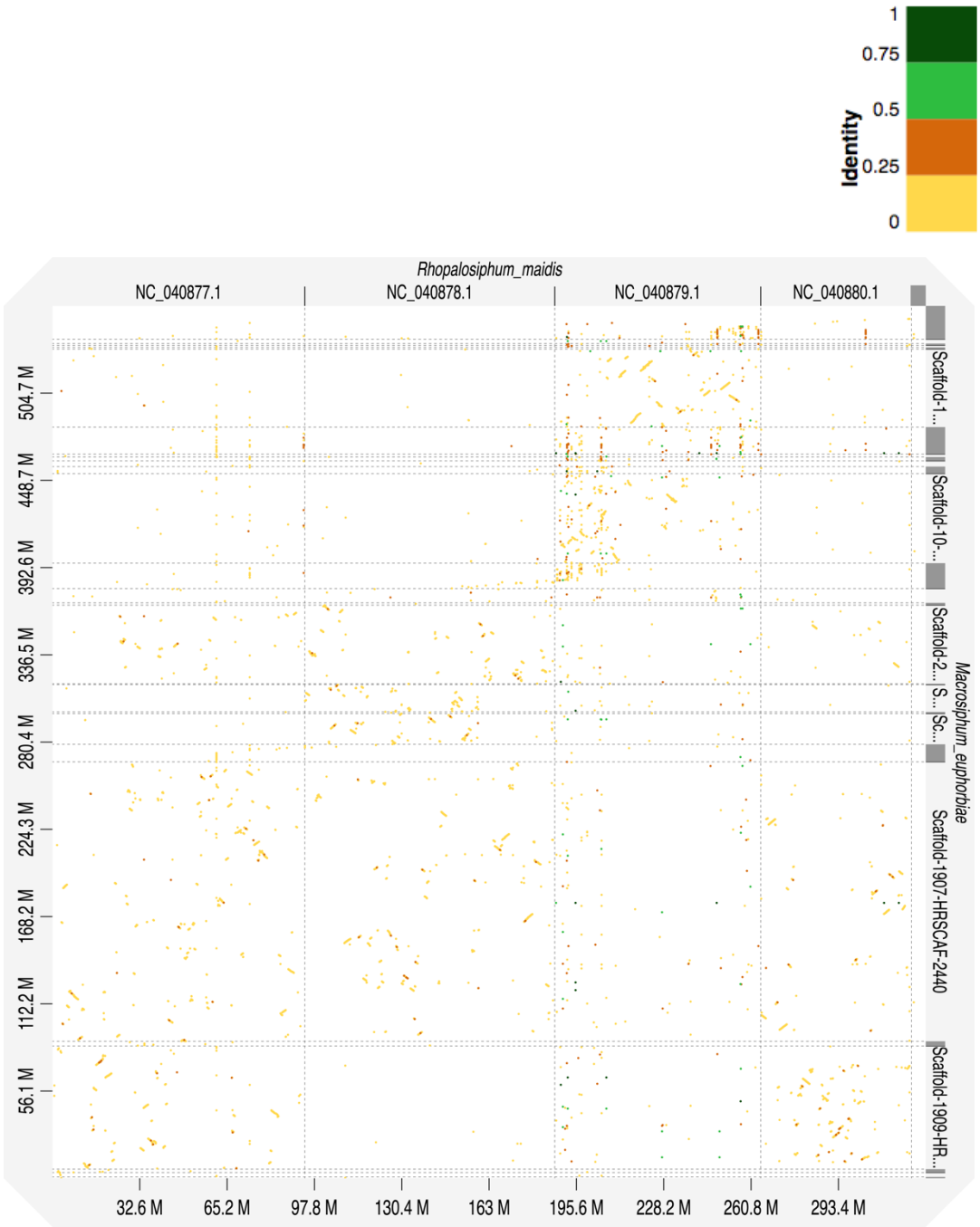
4.6 Supplementary Figures



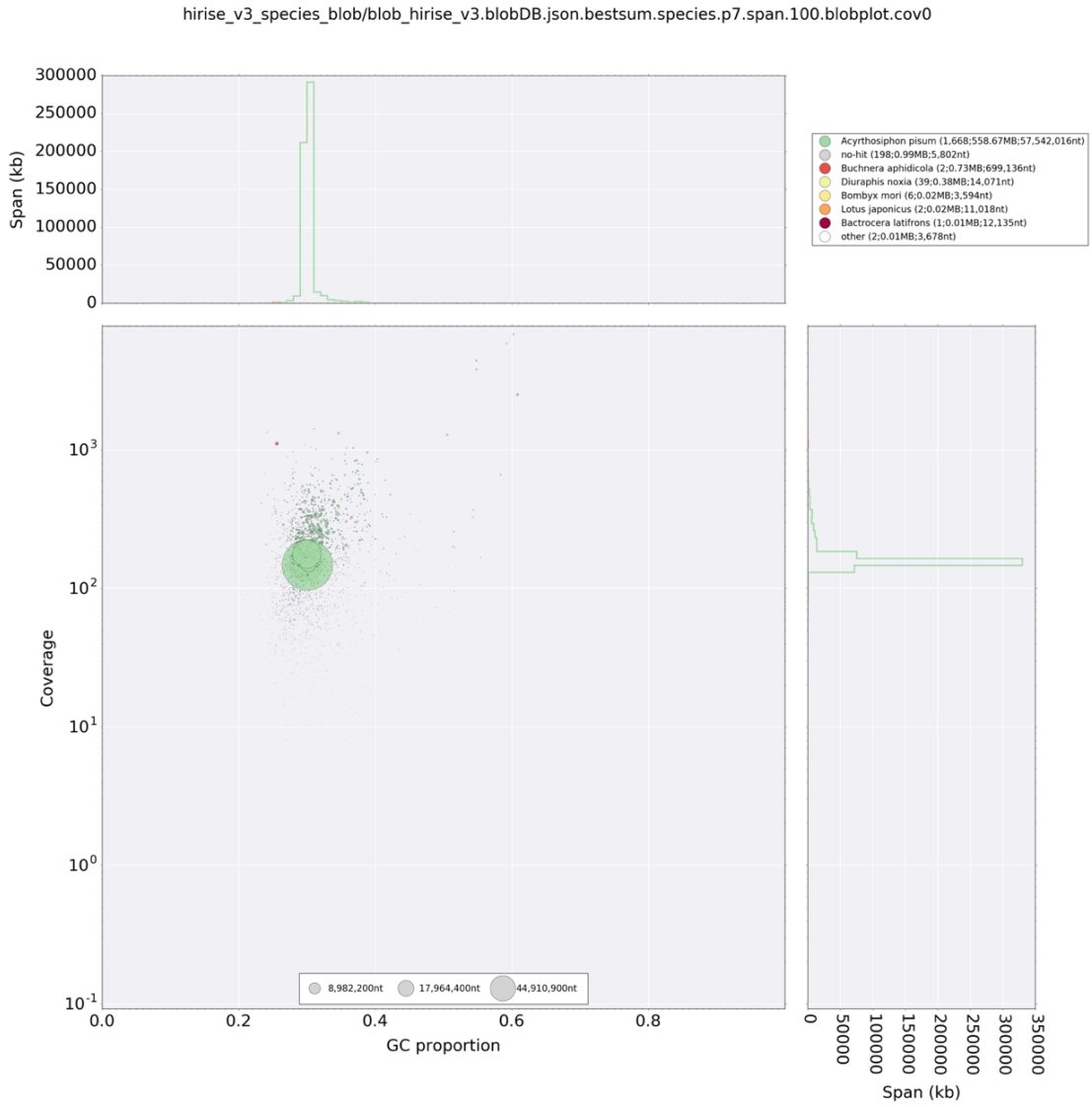
Supplementary figure 1. K-mer analysis of short reads for the genotype 1 line MW16/67. Genomescope predicts a genome size of 526 Mb, consistent with the predicted value in Wenger *et al.* (2017) of 530 Mb. GenomeScope also predicts more than 40% of the genome consists of repeat structures, as well divergence between haplotypes based on heterozygosity score. Low coverage in the sample results in poor resolution of haploid and diploid kmer peaks, which could be a result of *Buchnera* sequences present in the sample.



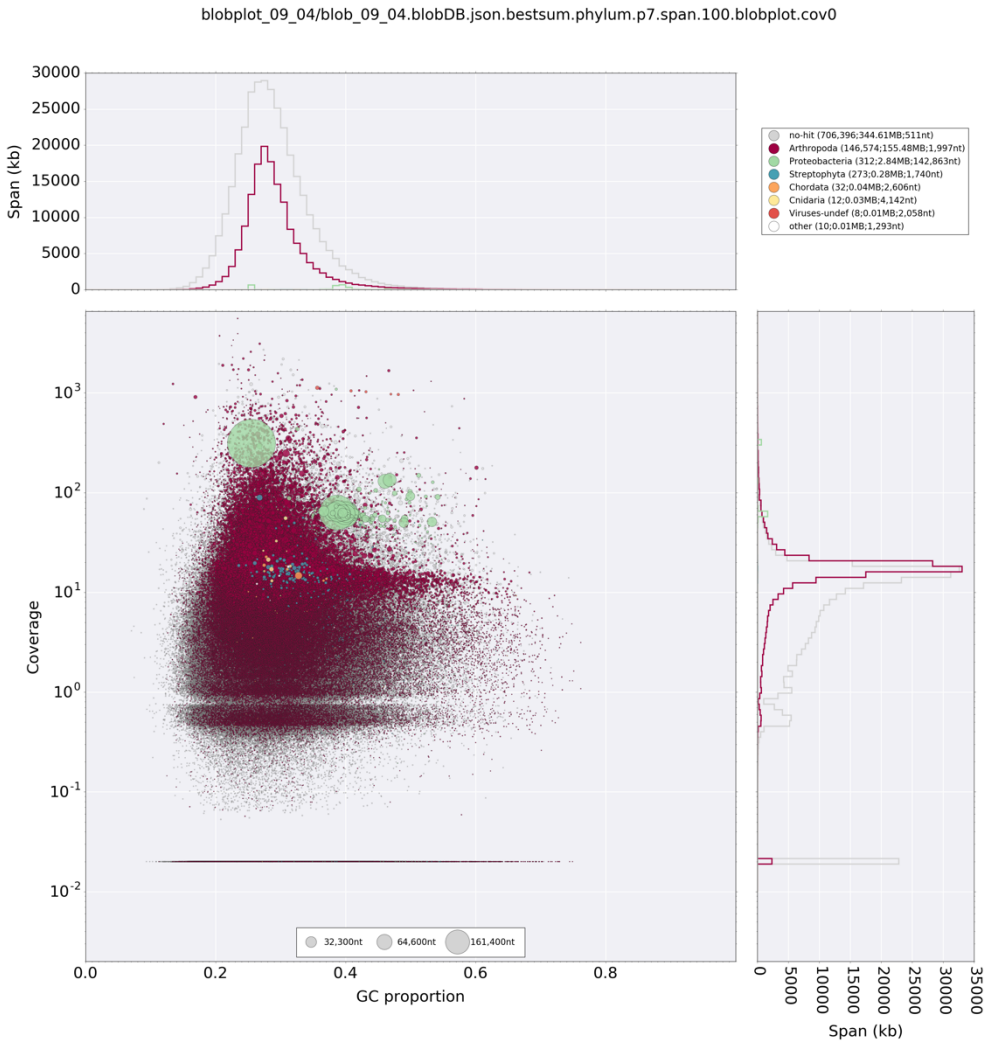
Supplementary figure 2. Whole genome alignment between *A. pisum* and *M. euphorbiae*. Alignment demonstrates conserved sequence between these two aphids. Colour of match highlights proportion identify between each matched sequence, where green indicates greater similarity, and yellow indicates less similarity.



Supplementary figure 3. Whole genome alignment between *R. maidis* and *M. euphorbiae*. Alignment demonstrates little sequence similarity of the Corn Leaf aphid to the Potato aphid, especially compared to the more closely related Pea aphid. Colour of match highlights proportion identify between each matched sequence, where green indicates greater similarity, and yellow indicates less similarity.



Supplementary figure 4. Blobplot generated from the HiRise assembly for *M. euphorbiae* (MW16/67). Little to no contamination is present in the assembly, with nearly all contigs and scaffolds showing BLAST hits to other aphid species.



Supplementary figure 5. Blobplot generated from the DISCOVAR assembly for *M. euphorbiae* line AA09/04. *B. aphidicola* was assembled into a single contig, along with its two plasmids pLeu and pTrp. Blobplot was generated at the phylum level.

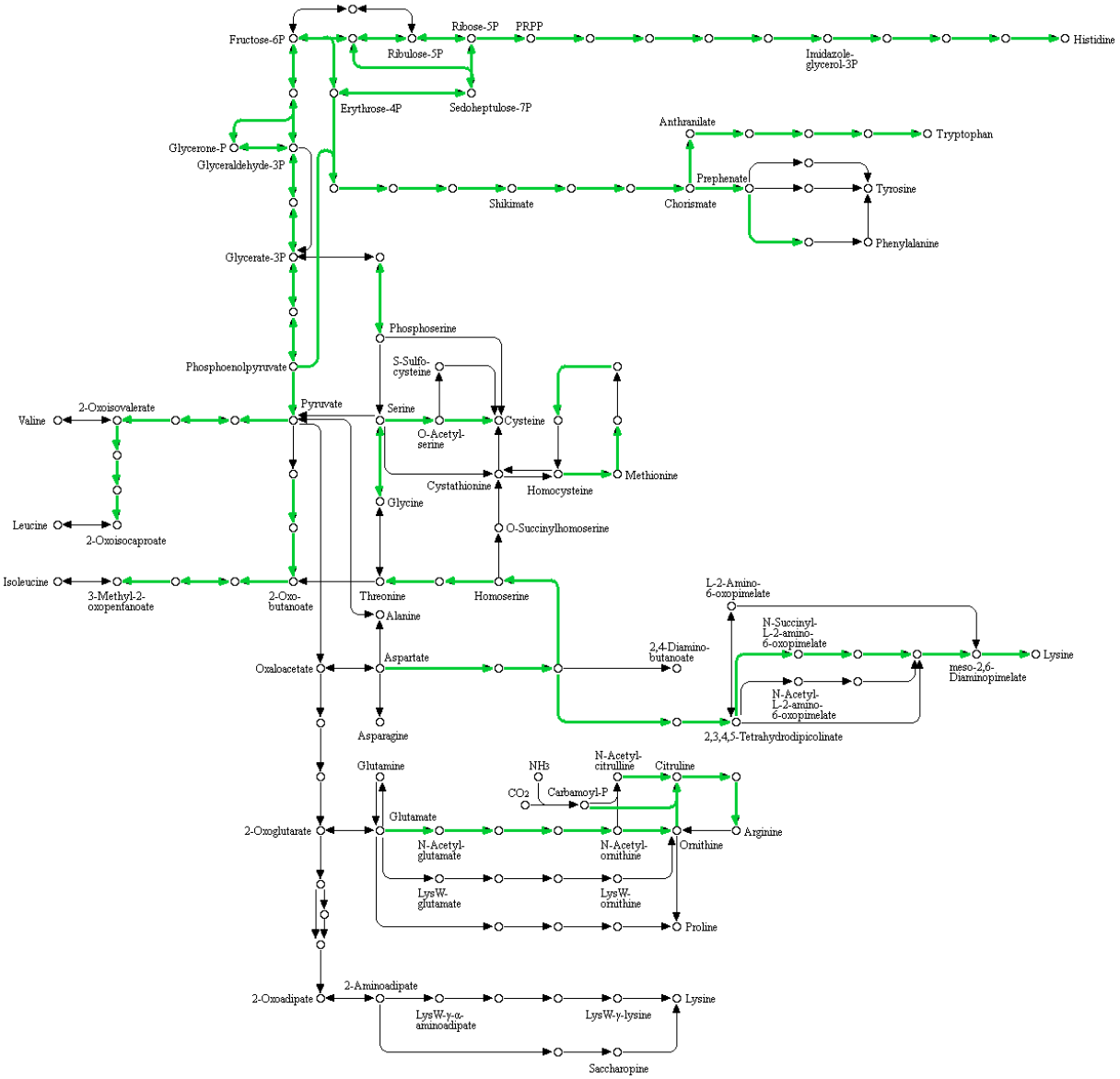
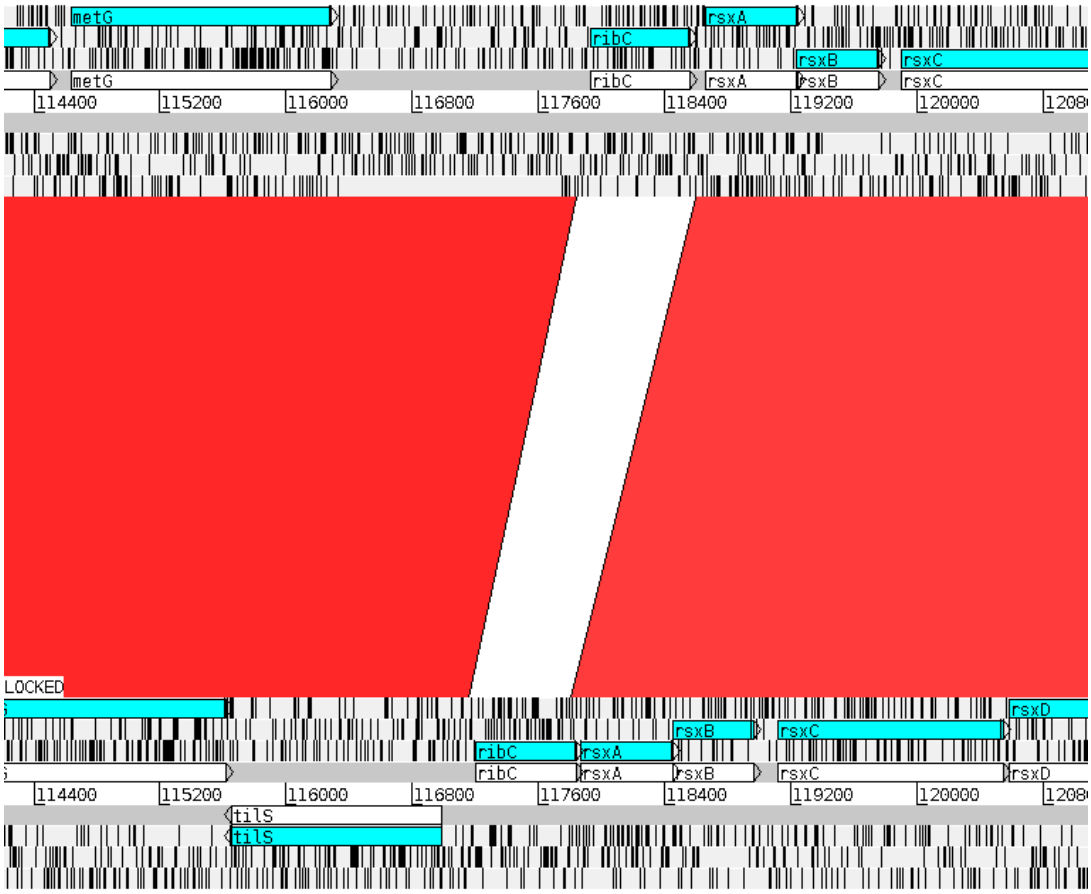
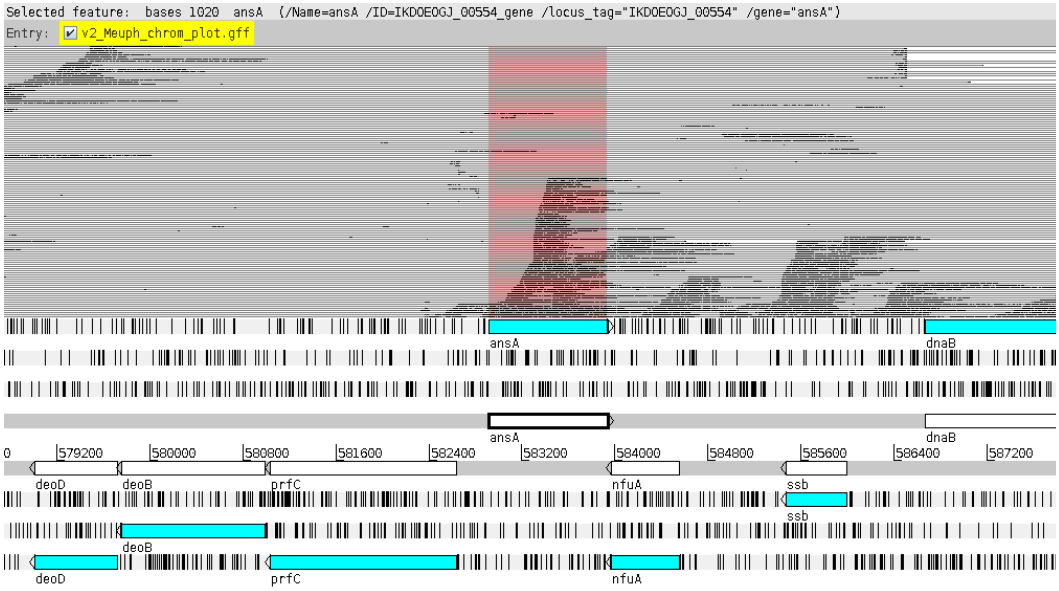


Figure 6. KEGG generated pathway of *Buchnera* genes required for biosynthesis of amino acids. Green arrows indicate the presence of genes identified on the *B. aphidicola* genome for *M. euphorbiae* generated in this study.

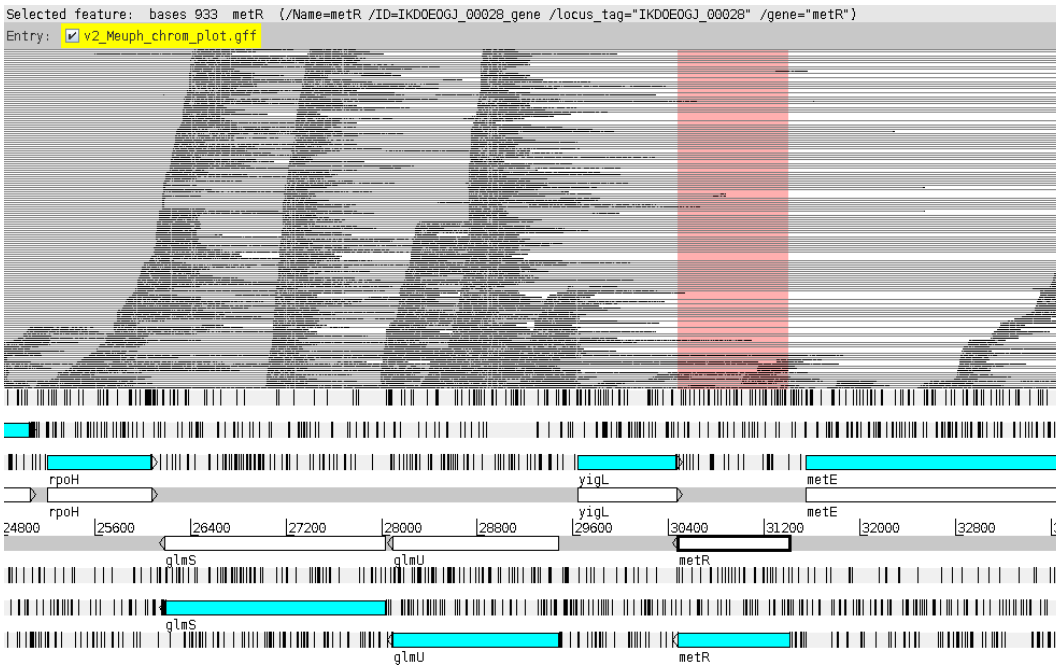


Supplementary figure 7. Lack of sequence similarity between riboflavin synthase C gene between *B. aphidicola* from *M. euphorbiae* (top sequence) and *A. pisum* (bottom sequence). Screen shot is from the ACT tool in Artemis. Red bars indicate sequence similarity between the two sequences. The lack of a handful of genes between *Buchnera* stains annotated as part of this study (e.g. see tilS gene above) could be the result of incorrect annotation or incorrect identification as a pseudogene. This implies further work is required to resolve and correct these issues.

A)



B)



Supplementary figure 8. A) Transcriptomic evidence for expression of *ansA* (asparaginase); B) *metR* (methionine biosynthesis regulator) shows little evidence for expression in *Buchnera aphidicola* from *M. euphorbiae*. Screen shots taken from Artemis with mapped ONT transcriptomic reads generated during this study mapped to the *B. aphidicola* reference. Vertical red bars indicate position of the gene, with dark horizontal bars being reads. Blue segments indicate genes, with arrows at wither end to denote orientation.

4.7 Other electronic resources

aphid_macrosiphum_euphorbiae_assembly.fasta
Final genome assembly of *M. euphorbiae* used in this study.
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Aphid_M_euphorbiae_protein.fasta
Protein sequences for draft annotation of *M. euphorbiae.*
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Aphid_M_euphorbiae_transcripts.fasta
Transcript sequences for draft annotation of *M. euphorbiae.*
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Aphid_M_euphorbiae.gtf
GTF file defining gene positions along the *M. euphorbiae* genome.
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Buchnera_genome.fasta
Final genome assembly of *B. aphidicola* used in this study.
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Buchnera_proteins.fasta
Protein sequences for draft annotation of *B. aphidicola*.
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

Buchnera_genes.gff
GFF file defining gene positions along the *M. euphorbiae* genome.
(http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/)

5. A multi-omic approach to assess clonal differences in genotypes of *M. euphorbiae*

Abstract

Macrosiphum euphorbiae in the UK is strongly suggested to persist asexually, as discussed during chapter 3. This results in the persistence of a handful of clonal lines, and the possible emergence of 'superclones', or clones with higher levels of fitness compared to others within the population. One example is the high level of resistance to parasitism by *Aphidius ervi* in a genotype of *M. euphorbiae*. The aim of this chapter is therefore to understand the link between Potato aphid genotype and fitness traits such as innate parasitism resistance by comparing genomic structural variations between genotypes followed by analysis of the metabolome profiles of genotypes 1 and 2 during *A. ervi* challenge. Asexuality was suggested from WGS and variant calling between 18 clonal lines and increased genetic distance of most clones based on PCA and *Fst*, with clonality also observed in aphid and *Buchnera* genomes within each aphid genotype. Clonal separation was also observed at the level of the metabolome between genotype 1 and 2 nymphs. Increased tissue riboflavin levels in genotype 1 identified via metabolomics also suggests a role for *Buchnera* in phenotypic variation. GATK variant calling, CNV analysis and assessment of paralogous genes under selection identified 4,613 candidate genes with variation specific to the parasitoid resistant genotype 1, of which 11 have suggested immune functionality. Metabolomics was unable to provide evidence for involvement of predicted immune pathways in phenotypic variation between genotypes, primarily due to a lack of annotated metabolites (n = 56). It does, however, demonstrate greater conversion of trehalose to glucose in the parasitism-resistant genotype, potentially for use as an energy source in mounting an immune response, similar to that observed in encapsulation response in other insects. The work here shows the degree of dissimilarity between *M. euphorbiae* genotypes, and indicates metabolic pathways that might be involved in overcoming parasitism by *A. ervi*.

5.1 Introduction

5.1.1 Effect of copy number variation on aphid phenotype

Phenotypic differences between aphid genotypes have been touched upon in previous chapters; for example, differences in host plant affinity and (Via, 1991), colour polymorphisms (Moran & Jarvik, 2010) in *Acyrthosiphon pisum*, *Myzus persicae* genotypes that demonstrate resistance to different insecticides (Bass *et al.,* 2014), and of course, parasitoid resistance in *Macrosiphum euphorbiae* (Clarke *et al.,* 2017). The Pea aphid also shows clonal variability in parasitoid resistance (Martinez *et al.,* 2014) however, the genetic basis underpinning this phenotype has still not yet been discovered.

Multiple routes to phenotypic changes can occur in aphids, such as gene sequence variants (Andrews *et al.,* 2004) and especially variations in gene copy number (Duvaux *et al.,* 2014). As previously highlighted when generating the genome of *M. euphorbiae* (Chapter 4), some genes linked to immune function exist in greater numbers in the aphid genome compared to other aphid species, and this pattern could exist for other gene categories within the Potato aphid genome leading to differences between Potato aphid genotypes. Copy number variations (CNVs) are large structural variants of genes or sets of genes and are the result of gene duplication or deletion events during genome evolution (McCarroll and Altshuler, 2007). Structural variation on this scale often leads to detrimental effects and often genetic diseases (Girirajan *et al.,* 2011). However, on some occasions CNVs can result in gene duplication or gene loss that has a positive effect on an organism and permits adaptive change (Kondrashov, 2012). These variations have previously been observed in other insects, especially in well-studied organisms such as *Drosophila melanogaster*, where X-linked genes and toxin-response genes were located on positively selected CNVs (Emerson *et al.,* 2008), and in the Mosquito *Culex pipiens*, which contains duplications of the *ace-1* locus and genes linked to insecticide resistance (Labbé *et al.,*2007). As previously stated, gene duplication events are suggested to be commonplace within the Pea aphid (IAGC,

2010). Further studies into *A. pisum* CNVs identify that they play a crucial role in speciation and adaptation to the local environment, specifically in gene families encoding olfactory and gustatory receptors (Duvaux *et al.,* 2014), which could be linked to defining Pea aphid biotype/host range (Via 1991; Ferrari *et al.,* 2006; Ferrari *et al.,* 2012).

5.1.2 Gene copy numbers with regards to the aphid immune system

Gene duplication is seen as a source of genetic variation for organism adaptation and evolution, with many duplications under positive selection when conferring a survival advantage (Cardoso-Moreira *et al.,* 2016; Haney *et al.,*2016; Van Zee *et al.,* 2016). Coupled with the observations that asexual organisms can show increased rates of deleterious and coding mutations (Henry *et al.,* 2012; Bast *et al.,* 2018), gene family expansion could give rise to multiple genic variants and associated changes in phenotype. With regards to the Potato aphid and obligate asexuality, gene gain/loss could be one of the main sources of variation that allows adaptation. Contrary to this, Ollivier *et al.* (2012) explored this theory, comparing rates of synonymous to non-synonymous mutation rates in gene orthologues between eight aphid species. They suggest increased missense mutation is not observed in aphids undergoing obligate parthenogenesis compared to aphids those that undergo cyclical parthenogenesis. The caveat here was their focus on 1:1 orthologues, which are generally more conserved due to their essential function.

Not only are gene duplications observed in immune related pathways within *A. pisum* (IAGC, 2010), aphids also show substantial gene loss, specifically in those linked to the immune deficiency (IMD) pathway (Gerardo *et al.,* 2010). In *Drosophila,* the IMD pathway is responsible for the expression of antimicrobial peptides (AMPs) in the haemolymph and responding to bacterial infection (Myllymäki *et al.,* 2014). However, the lack of observed peptidoglycan receptor proteins (PGRPs) for binding bacteria, and a lack of the IMD signalling cascade molecules *IMD, dFADD, Dredd, Kenny, Rel,* and *Kay* strongly indicates the pathway

is non-existent within pea aphids (Gerardo *et al.,* 2010). Selection for individuals lacking functioning IMD genes is thought to have been associated with infection by *Buchnera* endosymbiont, where IMD functions would otherwise prevent symbiosis (Douglas *et al.,* 2010). *Drosophila* cell lines that were artificially infected with *Buchnera* showed successful elimination of the symbiont, which further indicates that IMD gene loss is linked to symbiont establishment, rather than a suppression of the immune system by the endosymbiont (Douglas *et al.,* 2010).

While pathways like IMD show reduced functionality, others appear to have undergone expansion. A specific example is seen in dorsal, which shows and aphid-specific duplication (IAGC, 2010; Gerardo *et al.,* 2010). Dorsal is involved in both embryogenesis and morphology determination during development (Roth *et al.,* 1989; Hong *et al.,* 2008) and signalling within the toll pathway cascade, where in *D. melanogaster,* it plays important roles in the activation of cellular immune responses to Gram-positive bacteria or fungi (Valanne *et al.,* 2011). Specifically, the toll pathway is linked to induction of the prophenoloxidase (PPO) cascade, which results in melanization of foreign bodes, where invaders are encapsulated and smothered in melanin (Lemaitre & Hoffmann, 2007; An *et al.,* 2013). In aphids, however, the melanisation response to infection is suspected to be weak, perhaps to further facilitate the harbouring of secondary symbionts (Laughton *et al.,* 2011; Schmitz *et al.,* 2012).

Regarding *M. euphorbiae*, the previous chapter highlighted predicted immune gene paralogue numbers that vary between other aphid species, especially those related to the PPO cascade. Interestingly, only a single copy of proPO was identified, while conversely there is an increase in serpins, which are catalytic enzymes linked to induction of PPO and melanisation (Gorman & Paskewitz, 2001; An *et al.,* 2013; Ma *et al.,* 2019)

5.1.3 The potential of Metabolomics

Metabolomics is a recent and emerging field to study a 'snapshot' of an organism's metabolomic profile. This involves measuring concentrations of low molecular weight compounds (e.g. precursors, signalling molecules and end products of biochemical pathways (Oliver *et al.,* 1998; Viant *et al.,* 2017)) through proton nuclear magnetic resonance (1H NMR) and assessing chemical shift (ppm) of hydrogen atoms, where ppm differs depending on composition of the surround functional groups. The pattern of ppm peaks produced is distinctive to each compound and allows its annotation. What makes this technique so useful is that the process is non-destructive (samples can be measured repeatedly), requires little input and is relatively cheap compared to similar experiments such as large RNAseq studies.

While current metabolome databases are limited in number (Viant *et al.,* 2017), studies using metabolomics in insects are increasing. For example, Sato *et al.* (2015) were able to suggest compounds linked to foxglove aphid resistance in soybean plant strains, specifically those related to sulphur metabolism and methylation. Aside from aphids, metabolome investigations into the interaction between a tropical disease causing species of *Mycobacterium* and its mosquito vector highlight mechanisms of pathogen survival within the host (Hoxmeier *et al.,* 2015). Finally, *Drosophila* species have also been used in numerous metabolomics studies, due to their status as a model organism and the range of applications they can be used to study. *Drosophila* metabolomics studies are reviewed in Cox *et al.* (2017).

While whole metabolome identification may therefore be elusive for the Potato aphid, metabolomics studies may still provide useful information about the biology of *M. euphorbiae*. For example, most amino acids can be identified readily using 1H NMR, and this group of compounds is integral to many metabolic pathways (Gu *et al.,* 2015). The movement and sharing of amino acids and other small compounds between the host aphid and *Buchnera* symbiont has been

characterized in detail, such as the provision of tryptophan from *B. aphidicola* to the aphid (Douglas & Prosser, 1992); the movement of glutamine from the aphid to *Buchnera* via specific transporters in the bacteriocyte cell membrane (Price *et al.,* 2014); and riboflavin provision from *Buchnera* for aphid nymph development (Nakabachi & Ishikawa, 1999). SNP variants have previously been identified in *ribF* that result in reduced efficiency of the encoded riboflavin kinase and a build-up in riboflavin in *E. coli* (Wang *et al.*, 2015). Therefore, any potential variation in *Buchnera* could have knock-on effects on the aphid host.

Regarding immunity, fluctuations in amino acid concentrations have also been observed in insect immune responses, such as the role of glutamate in *D. melanogaster* phagocytosis (Gonzalez *et al.,* 2013) and levels of phenylalanine and tyrosine during parasite melanisation within the mosquito *Anopheles gambiae,* where phenylalanine conversion to tyrosine is upregulated for use in melanin biosynthesis (Fuchs *et al.,* 2014). Conversion of sugars as an energy source could also highlight specific immune pathways. Trehalose is more often linked to flight in insects, where trehalase breaks down trehalose to glucose for energy in wing-beating (Becker *et al.,* 1996). However, trehalose has also been implicated as important factor in immunological responses. For example, the silencing of a trehalose-6-phosphate synthase (used for trehalose synthesis) in the housefly *Musca domestica* results in reduced insect survival in response to bacterial challenge (Zhang *et al.,* 2019), where normally, trehalose synthesis increases after exposure to the pathogen. Glucose has previously been reported as important in cellular immune responses in insects (Dolezal *et al.,* 2019). Specifically, in *Drosophila* it is required for the maintenance of lamellocytes, specialised haemocytes used in encapsulation and melanisation in response to the parasitoid larvae of *Leptopiliana boulardi* (Bagjar *et al.,* 2015). However, aphids lack specialised haemocytes for such a function (Laughton *et al.,* 2011).

Studies have previously suggested a role for sequence variation in *Buchnera* to have an effect on the aphid/*Buchnera* holobiont (Dunbar *et al.,*2007). Regarding nutrient provision, Jiang *et al.* (2013) reported variation in the threonine synthase

gene*, thrC*, involved in essential amino acid threonine generation, between *Buchnera* strains belonging to different clones of *M. persicae*. Tryptophan provision by *B. aphidicola* has also been shown to differ based on aphid clonal genotype (Birkle *et al.* 2002), however this is believed to be a dynamic process involving variation in both aphid and endosymbiont genomes.

5.1.4 Specialised parasitoid mechanisms for survival within a host

Embryonic cells of the wasp larvae develop in specialised cells known as teratocytes, which hijack host metabolism and siphon amino acids away from aphid ovarioles to the developing parasitoid (Falabella *et al.,* 2000; Strand, 2014). Amino acid titres in the aphid are, therefore, an important factor for developing parasitoid larvae within aphid hosts, as amino acid levels change dramatically in response to parasitism (Cloutier, 1986). In *A. pisum*, parasitism by *A. ervi* induces a hypertyrosinemic environment as a result of increased phenylalanine transfer to *Buchnera,* where phenylalanine-to-tyrosine conversion occurs (Rahbé *et al.,* 2002). Teratocyte function is further verified in Li *et al.* (2002), where failed teratocyte development is detrimental to parasitoid survival, and could be a source of innate resistance to *A. ervi.*

Successful *A. ervi* parasitism of aphids is not reliant on teratocyte cells alone, however. Parasitism is further supported through co-injection of parasitoid venom during oviposition, which may harbour component such as polydnaviruses (PDV) that can influence host immunity and optimize growth conditions for parasitoid larvae (Burke & Strand, 2012). *A. ervi* is not believed to utilize PDV, but instead relies on proteins that cause host castration (Digilio *et al.,*2000; Colinet *et al.,* 2014). *A. ervi* venom contains three putative gamma-glutamyl transpeptidases (γ-GTs) that induce castration through the build-up of reactive oxygen species (ROS) in aphid ovarioles and results in their apoptosis (Colinet *et al.,* 2014). However, ROS build up can be avoided through their conjugation to glutathione (GSH), performed by glutathione-S-transferases (GSTs). This is commonly observed as a method of resistance to xenobiotics and insecticides, where the conjugation of

GSH to toxic exogenous compounds allows easier solubilisation and subsequent removal from the host (Constant *et al.,* 2004; Enayati *et al.,* 2005; Olivier & Brooke, 2016). Delta-class GSTs specifically are linked to recent emergence of insecticide resistance, with evidence suggesting they are a rapidly evolving gene family (Ranson & Hemingway, 2005). GSTs may also counteract build-up of endogenous ROS during times of cell stress, such as after exposure to extreme cold (Lalouette *et al.,* 2011) and generally during elevated immune responses (Pedra *et al.,* 2003).

5.1.5 Aims and objectives

Here, I aim to take a multi-omic approach to characterise the differences between genotypes of *M. euphorbiae*, and potentially link these variations to genotypic variation in parasitoid resistance. The work here attempts to investigate the following:

1. Assess genetic relatedness between six prevalent genotypes of *M. euphorbiae* and their respective *B. aphidicola* genomes through whole genome sequencing and variant calling against the previously generated genotype 1 (parasitoid resistant) reference. Genic variants specific to genotype 1 (those which are found in all susceptible genotypes) will provide candidate genes for parasitoid resistance.
2. Assess paralogous genes for positive selection via *dN/dS* analysis, identifying genes which accumulate and maintain variants that change amino acid sequence of a protein, and potentially its functional efficiency. Gene orthologue prediction between aphid species provided information about paralogous genes within *M. euphorbiae*, which have potential effects on aphid phenotype. Gene gain and gene loss in genotype 1 will also be studied through CNV analysis. Again, genes identified here will be candidates for parasitoid resistance.
3. Parasitism resistance is observed in *M. euphorbiae* as early as second instar, therefore metabolomes of four-day old Potato aphid nymphs (genotypes 1 and 2 only) are studied in response to challenge or without

prior challenge by the parasitoid *A. ervi*, assessing links between metabolite patterns. Changes in compounds linked to aphid*/Buchnera* mutualism and specific immune response pathways such as glutathione metabolism (oxidative stress) and phenoloxidase pathways as well as utilising of sugars will be a specific focus.

Demonstrating contrasts between Potato aphid genotypes in their genetic sequence, and linking these to differences in metabolomes, could provide clues to uncover the basis for innate parasitoid resistance in *M. euphorbiae*.

5.2 Methods

5.2.1 Genomic DNA extraction and preparation of short read Illumina libraries

Genomic DNA extraction was performed as outlined in section 2.2. Library preparation and sequencing for six *A. ervi* resistant genotype 1 lines is outlined in section 4.2.2. A further 10 TruSeq PCR free libraries were generated consisting of parasitoid susceptible genotypes (table 1) for sequencing on the Illumina 4000, as well as two TruSeq PCR free libraries for samples MW16/48 and AK13/30 (2x300 bp) for the Illumina HiSeq 2500 (rapid run mode). All reads were trimmed for Illumina adapters using Cutadapt (v1.2.1) (Martin, 2011) followed by further trimming and quality filtering with Sickle ("-q 20" and "-l 20" enabled) (v1.200) (Joshi & Fass, 2011).

5.2.2 Read mapping to the M. euphorbiae reference genome

Reads were first mapped to the MW16/67 genotype 1 (see chapter 4) using bwa mem (v0.7.5a-r405) (Li & Durban, 2009) with each read set given a unique read-group sample name based on clonal line (table 1). Read duplicates that were potentially introduced during library preparation were filtered out using MarkDuplicates as part of picard tools (v2.8.2) (Broad Institute, 2019). Whole genome coverage was assessed using 'DepthOfCoverage' as part of the GATK package (v3.7) (Mckenna *et al.,* 2010).

Table 1. Clonal lines of *M. euphorbiae* used in this study. '*' denotes lines collected prior to the study.

Clonal Line	*Genotype*	*Symbiont status*	*Parasitoid susceptible*
MW16/67	1	-	Yes
AA09/03*	1	*H. defensa*	Yes
AA09/04*	1	*H. defensa + APSE*	Yes
R. res*	1	*H. defensa*	Yes
AK13/27	1	-	Yes
RB15/10*	1	*H. defensa + APSE*	Yes
AK13/30	2	*H. defensa + APSE*	No
MW16/38	2	*R. insecticola/H. defensa + APSE*	No
MW16/48	2	-	No
MW17/01	2	*H. defensa + APSE*	No
MW17/23	2	*H. defensa + APSE*	No
MW17/33	2	*R. insecticola/H. defensa + APSE*	No
MW16/52	3	-	No
MW16/88	3	*H. defensa + APSE*	No
MW16/40	6	-	No
MW17/31	6	*R. insecticola/H. defensa + APSE*	No
MW16/113	7	-	No
MW16/98	p7 (pink)	*H. defensa + APSE*	No

5.2.3 Variant calling between aphid genotypes

Calling genetic variants between aphid genotypes provides one method for determining genetic differences that might underlie parasitoid resistance as well as permit a higher resolution method for genotyping, compared to microsatellite markers used in chapter 3. Single nucleotide polymorphisms (SNPs) and insertion/deletions (INDELs) will occur in genomes over-time. Variants may occur in gene encoding regions and could affect gene expression or efficiency of the encoded product. The specificity of these variants to whole genotypes instead of individual clonal lines should further support an obligate parthenogenetic lifestyle in *M. euphorbiae*.

GATK (v3.7) was used for variant identification between Potato aphid clonal lines (Mckenna *et al.,* 2010). Mapped reads were re-aligned around INDELs using the two GATK tools 'RealignerTargetCreator' and 'IndelRealigner'. 'HaplotypeCaller' was used to generate variant calls for each clonal line mapped to the genotype 1 reference (--GVCF" enabled), followed by joint genotyping and combing of individual sample variant calls using 'GenotypeGVCFs'. Variants were separated into separate SNP INDEL vcf files and filtered using 'VariantFiltration', using the following filter expressions; "QUAL < 0 || MQ < 40.00 || SOR > 4.000 || QD < 2.00 || FS > 60.000 || MQRankSum < -20.000 || ReadPosRankSum < -10.000 || ReadPosRankSum > 10.000" for SNPs, and "QUAL < 0 || MQ < 40.00 || SOR > 10.000 || QD < 2.00 || FS > 60.000 || MQRankSum < -20.000 || ReadPosRankSum < -10.000 || ReadPosRankSum > 10.000" for INDELs. Variants were further filtered on depth of coverage (DP) with the expression "DP < 6 || DP > 50" and the option "--setFilteredGtToNocall" enabled.

5.2.4 Assessing genetic relationship between Potato aphid genotypes

Clustering aphid SNPs assesses the divergence between potato aphid genotypes. A subset of SNPs was generated with a minimum of 5000 bp between each-other using vcftools ("--thin 5000") (v0.1.13) (Danecek *et al.,* 2011). Genotype clustering and MDS tables were generated through PLINK (v1.90p) (Purcell *et al.,* 2007) ("--noweb" and "--allow-no-sex" enabled). The multidimensional scaling (MDS) table contains 10 principal components (PC) based on genetic distance between individuals using SNPs; the first two principal components, PC1 and PC2, explaining 58.0% of the variation were used for PCA. PCA variation for each component was calculated as a percentage of the sum of all eigenvalues in the generated "plink.eigenval" output for each individual component. PCA plots were generated in R (v3.4.4) (R core team, 2015).

Genetic similarity between aphid genotypes was also assessed via Weir and Cockerham's fixation index (*Fst*). Pairwise *Fst* values for each combination of genotypes were calculated using vcftools (Danecek *et al.,* 2011). Vcftools tools

was provided with the final set of SNP variants as well lists of samples specific to each genotype assessed.

5.2.5 Generating genotype 1 specific gene variants

To focus analysis on variants observed in the parasitoid resistant genotype 1, SNP and INDEL variant positions shared between all susceptible clonal lines only were selected using a custom script provided by CGR (University of Liverpool). This specifically focused on variants that could affect gene function. Therefore, those that passed filtering were annotated for putative variant effect using snpEff (v4.3t) (Cingolani *et al.,* 2012), followed by further selection of variants using SnpSift (SnpEff package) with the filter "ANN[*].EFFECT has 'missense_variant'" for SNPs and "ANN[*].EFFECT has 'frameshift_variant'" for INDELs.

5.2.6 Variant calling between genotypes of the Buchnera symbiont

Short reads from 18 clonal lines of *M. euphorbiae* were provided to snippy (v4.3.6) (Seemann, 2015) to call SNP and INDEL variants simultaneously against the *Buchnera* reference from AA09/04, as generated in section 4.3.7. Briefly, the snippy pipeline involves read mapping with bwa (Li & Durban, 2009) followed by variant calling using freebayes (Garrison & Marth, 2012) and variant annotation using snpEff (Cingolani *et al.,* 2012). Initial testing of the pipeline was slow due to the time taken in mapping excessive coverage across the small *Buchnera* symbiont genome (645 Kb) (*Buchnera* high coverage is also visualised in sup. figure 4.6.5) Reads were subsampled within snippy to approximately 100X coverage of all reads ("--subsample 0.1" enabled) to speed up alignment and variant calling. Phylogenetic trees from SNPs were generated using the Gubbins package (v2.3.4) (Croucher *et al.,* 2015) and FastTree (v2.1.10) (Price *et al.,* 2010). Phylogenetic trees were visualised in FigTree (v1.4.3) (Rambaut, 2009).

5.2.7 Assessing presence of copy number variation between parasitism-resistant and -susceptible Potato aphid genotypes

Copy number variations (CNV) have previously been identified as a source of phenotypic variance in aphids (Duvaux, 2014). To identify genes which are affected by CNV in genotype 1 of *M. euphorbiae*, a depth-of-coverage based approach was taken. CNV analysis was performed using control-FREEC (Boeva *et al.,* 2011). De-duplicated and sorted mapped Illumina reads generated during variant analysis were provided as a merged bam file of either all resistant clones, or susceptible clones (read counts for samples MW16/48 and AK13/30 were not used in this analysis due to differences in library preparation compared with the other samples. The sample MW16/52 was also not included due to low coverage). CNV was measured over sliding windows of 10,000 bp. The option 'telocentromeric length' was reduced from the default of 50,000 to 25,000 as outlined in the control-FREEC manual for non-human genomes. 'minExpectedGC' and 'maxExpectedGC' options were changed from default to 0.25 and 0.35 respectively, to be more in line with expected GC content of the aphid genome. The control-FREEC output '*_ratios.txt' was parsed to select all predicted CNV regions that overlapped with gene predictions. CNV plots were generated for selected *M. euphorbiae* scaffolds using 'makeGraph.R', as part of the control-FREEC package.

5.2.8 dN/dS analysis of genotype 1 paralogues genes single copy orthologues to the Pea aphid

Gene family expansion and gene duplication is commonly observed in aphid species (IAGC, 2010; Mathers, 2017). Therefore, paralogous gene families were assessed for positive selection pressure. Paralogues within the genotype 1 Potato aphid were identified from previously performed Orthofinder analysis (performed as part of section 4.3.6) (v7.407) (Emms & Kelly, 2018). The longest of each Paralogue group was retained as well as its closest BLASTp hit ("-evalue 1e-10" enabled), using similar parameters outlined in Mathers *et al.* (2017) and Li *et al.* (2019); i.e. those with minimum protein length alignment of 150 amino acids and

a minimum of 30% shared identity. Possible large evolutionary distances between paralogues, as well as shared identical amino acid sequences in functional domains during global alignment (end-to-end) permit these the use of such a low identify threshold (Pearson, 2013). Paralogue pairs were aligned using Mafft (v7.407) (Nakamura *et al.,* 2018). Protein alignments were converted to codon alignments using PAL2NAL (v14.0) (Suyama *et al.,* 2006). Ratios of non-synonymous (*dN*) mutations (those that change the amino acid sequence in a protein) to synonymous (*dS*) mutations (those that do not change the amino acid sequence due to codon degeneracy) were calculated for each paralogue pair using the codeml wrapper as part of the paml package (v4.9h) (Yang, 2007) and a pairwise maximum-likelihood method. Those with $dS < 0.01$ or $dS > 2$ were removed from analysis, where *dS* being too low indicates very similar sequences and would provide an unreliable *dN/dS* result, as would *dS* being too high, where nucleotide sequences would be too dissimilar. Those with $dN/dS > 1$ were considered to be under positive selection pressure, where genes maintained a greater number of non-synonymous mutations that synonymous mutations. Predicted function of positively selected proteins was generated through InterProScan (v5.30-69) (Jones *et al.,* 2014). A similar method was applied to studying one-to-one orthologues between the *A. pisum* and *M. euphorbiae* reference genomes (see section 2.4 regarding *A. pisum* data sources). Single copy orthologues were previously identified in Orthofinder analysis (performed as part of section 4.3.6). Potato aphid and Pea aphid orthologues were then analysed in the same fashion as paralogue pairs as previously described. *dN* and *dS* were plotted against each other using R (v3.1.1) (R core team, 2015), as were boxplots demonstrating *dN/dS* ranges between orthologues and paralogues. Comparing *dN/dS* values can indicate whether paralogous genes are under positive selection compared to single copy orthologues that are more likely to be conserved (i.e. do not accumulate missense mutation).

5.2.9 Parasitoid assays for metabolomics analyses

Potato aphids were cultured and maintained as described previously in general methods (section 2.1). Apterous adult aphids were transferred to a freshly-excised leaf in a culture cup and allowed to deposit nymphs overnight. A minimum of 60 aphid nymphs were set aside for parasitoid attack and no-attack controls from aphid clonal lines MW16/67 and MW16/48, belonging to genotypes 1 and 2 respectively. Parasitoid assays were carried out as previously described (section 2.3) using 4-day old nymphs.

Nymphs were flash frozen in liquid nitrogen immediately prior to attack (i.e. at 0 hours before attack), and at 6 and 12 hours after the parasitoid assays were initiated for attacked nymphs and no-attack controls; samples were stored at -80 °C prior to extraction. For each test condition and time-point, nine replicate nymphs were generated. Due to known inefficiencies in *A. ervi* parasitism (no 100% success rate (McLean & Godfray, 2017), also see 3.3.4)), it was decided as little material would go into each replicate as possible. This was also helpful for experimental logistics, to reduce the time between first and last ovipositions and keep samples as identical as possible.

5.2.10 Sample processing for 1H NMR

Frozen samples were submerged in 500 μL of ice-cold solvent solution (50% HPLC grade acetonitrile; 50% ultrapure water). Samples were sonicated in three 30 second bursts, with 30 seconds between each sonication. Samples were vortexed briefly and centrifuged at 21,000 *g* for 5 minutes at 4 °C. The supernatant was lyophilized, then stored at -80 °C. *

**Some samples processed by Lianne Davies during my field work in Dundee.*

To prepare samples for NMR acquisition, metabolites stored at -80 °C were re-suspended in 200 μL of 1 mM sodium phosphate buffer (pH 7.4 in 100 % D_2O) and 0.2 μL of 100 mM Trimethylsilylpropanoic acid (TSP). Samples were vortexed briefly and centrifuged at room temperature for 2 minutes at 12,000 *g.* Finally, approximately 200 μL of supernatant was transferred to 2 mm NMR glass tubes using a glass Pasteur pipette. 1H NMR metabolite measurement was performed by the NMR Centre at the University of Liverpool.

It was vital to assess the quality of the metabolite peaks detected. Raw spectra were visually inspected using TopSpin (v3.5 patch level 6) (Bruker, Massachusetts, USA). The TSP peak at 0.00 ppm acts an internal standard, where the peak size can be compared between samples run within the same batch. Within each batch, the average and standard deviation was calculated for the TSP peak. Those that lie outside the standard deviation would be removed from further analysis.

5.2.11 Metabolite identification

Identification of metabolites was performed by the NMR centre at the University of Liverpool. Using the best spectra from a preliminary dataset (generated prior to this study) with the best signal-to-noise ratio belonging to MW17/04 (winged, genotype 3), a pattern file of identified compounds was manually curated. The ppm for each metabolite peak was measured (relative to the 0.00 TSP peak), with combinations of specific peaks relating to specific compounds. Annotation was carried out using Chenomx (Chenomx, Alberta, Canada) which contains a library of approximately 360 characterised human metabolites to which the aphid spectra was compared. Metabolite data for *Daphnia* and a Mosquito species generated by the NMR centre (unreleased data) was also used to aid annotation. *

**Manual annotation of the metabolome was performed by Dr Marie Phelan.*

5.2.12 Metabolite normalisation and scaling

Metabolite concentrations per extraction may differ between samples. Pre-processing of the data can account for these differences to allow comparative analyses between identified peaks. Normalization allows us to account and remove unwanted variation within a dataset (e.g. variation in absolute metabolite concentration between samples), while scaling applies filters to a normalized dataset to help accurately pull apart potentially significant from non-significant metabolites (Livera *et al.,* 2016). In-house R scripts provided from the NMR Centre were used to normalize and scale the data. Metabolite peaks were normalised to total area below the spectrum (percentage of peak abundance divided by the sum of total peak abundance of all metabolites within a sample). Scaling was performed through mean centering and pareto scaling.

5.2.13 Statistical analyses to identify significantly different metabolite peaks

Normalised and scaled metabolite peak sizes (corresponding to metabolite concentration) were uploaded to the online MetaboAnalyst web portal (v4.0) (Chong *et al.,* 2018) for statistical testing and plot generation. The following analysis focused on time points of 0 hours and 12 hours post-attack by *A. ervi*. Variation between treatment group metabolomes was visualised using PCA. For four-day old nymphs, metabolite peaks were identified that differed significantly between the genotypes 1 and 2 using *t-test* and false discovery rate (FDR) correction as a method to remove false positives (Benjamini & Hochberg, 1995), with further differences were highlighted by assessing log2 fold change (log2FC) of each compound. For treatment groups 12 hours post-attack, significantly different metabolite peaks between aphid genotypes 1 and 2 and/or between *A. ervi* challenged/unchallenged aphids were identified through two-way ANOVA and FDR correction. Peaks were considered significant if their FDR-adjusted *P* value was below 0.05. Plots pertaining to metabolite data analysis was generated using the MetaboAnalyst webportal. A Venn diagram of significant metabolite peaks was generated using the online Venny tool (Oliveros, 2007).

5.3 Results

5.3.1. Sequencing results

Whole genome sequencing of 18 different Potato aphid lines was generated over three separate runs over the course of the project. Sequencing of six genotype 1 clonal lines yielded a total of 82.8 million paired-end reads and consisted of 121X coverage. Sequencing of 10 susceptible genotypes yielded a slightly lower coverage in total, generating 74.1 million paired-end reads and 110X coverage. Finally, two libraries sequenced on the HiSeq 2500 in rapid run mode yielded 13.6 million paired-end reads and 19X coverage. The average whole genome coverage per sample is approximately 14X.

5.3.2 Clonality in M. euphorbiae genotypes is observed at both the level of aphid host and symbiont

As previously observed through microsatellite analysis of field-collected aphid samples, most (if not all) Potato aphids found in the UK can be classed into clear genotype categories (section 3.3.1). Using a small number of microsatellite positions provides enough evidence to deduce these genotypes, although strict clonality is visualised at a higher resolution through variant analysis and clustering of individuals based on SNPs (figure 1). Aphids of genotype 1 show very high genetic similarity based on the overlapping of PCA data points. Aphids belonging to the most common genotype, genotype 2, exhibit some SNP divergence, based on slight separation of points within the cluster (figure 1). High values for *Fst* estimates between each pair of genotypes also demonstrate the amount of divergence between genotypes, and further support obligate parthenogenesis in potato aphids and the presence of a handful of clonal genotypes (table 2) *Fst* in clonal species would be higher compared to species that reproduce sexually due to a lack of genetic recombination (Lambertini *et al.,* 2010; Yukilevich *et al.,* 2018). Genotypes 2 and p7 appear the most closely related based on lower *Fst* and PCA clustering, and could indicate a recent evolutionary divergence compared to other genotypes. Variant calling with GATK generated 17,815,413 SNPs and 3,371,150

INDELs across all other genotypes against a genotype 1 reference (MW16/67). Of all variants, 341,564 SNPs and 53,658 INDELs are shared between all susceptible genotypes and are not present in genotype 1 lines; of these, 3,305 SNPs (affecting 2,043 predicted genes) are missense variants and 244 INDELs cause frameshifts (affecting 197 predicted genes).

Genic variants are linked to multiple immune pathways, and summarized in table 3. Most genotype 1 specific variants affect jak/stat related proteins (n = 7) with domeless 2 harbouring the most missense variants (n = 6). There is also increased variation in four toll-pathway linked genes, implicating melanisation as a pathway of interest. There appears to be no variants linked to what remains of IMD/JNK pathways in the aphid genome.

Table 2. Estimates of pair-wise *Fst* between Potato aphid genotypes. Higher *Fst* values indicate lack of shared sequence between compared aphid genotypes. '*' Unable to calculate *Fst* between two individual samples.

	1	***2***	***3***	***6***	***7***	***p7***
1	-	-	-	-	-	-
2	0.32	-	-	-	-	-
3	0.42	0.32	-	-	-	-
6	0.39	0.26	0.21	-	-	-
7	0.46	0.33	0.32	0.35	-	-
p7	0.39	0.11	0.34	0.31	NA*	-

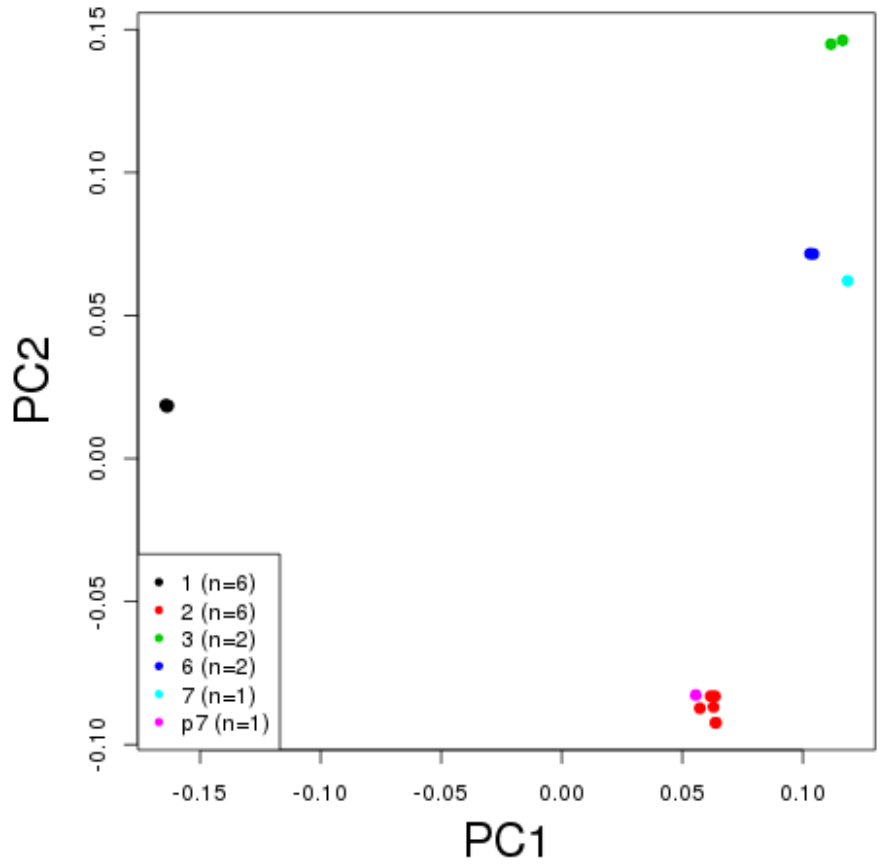
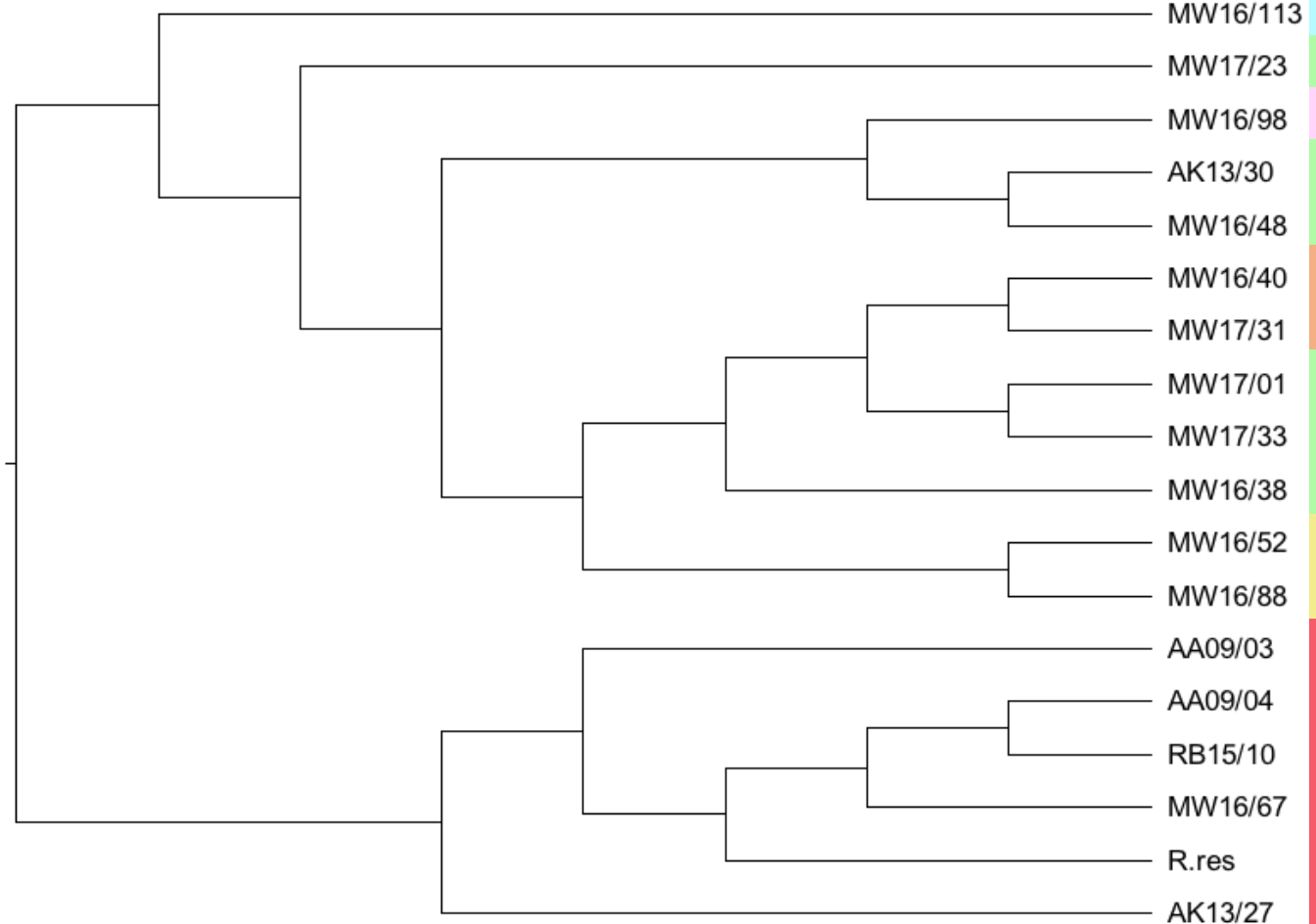


Figure 1. PCA analysis of SNP variants between six genotypes of *M. euphorbiae* identified by microsatellite analysis. Attribute loadings are shown on the first two principal components, which explain 35.2% and 22.8% of the variation in the data for PC1 and PC2 respectively. Little overlap between genotypes (especially with genotype 1) suggest strict clonality within aphid genotypes, supporting the lack of recombination between aphid lines and cyclical parthenogenesis. PCA was generated using a subset of 104,614 SNPs.

Table 3. Shared variants between parasitoid susceptible genotypes against the resistant genotype 1 within predicted immune-functioning genes. Reference refers to the genotype 1 (*A. ervi* resistant) Potato aphid genome. Variants are reported as genome variants, rather than variants of cDNA.

gene ID	*Scaffold*	*predicted immune function*	*loci*	*reference/variant*
g20675	Scaffold-10	domeless 2 (jak/stat)	36517786	G => C, His => Asp
			36517789	C => T, Val => Ile
			36517791	G => A, Thr => Met
			36517800	C => T, Arg => His
			36520673	T => A, Thr => Ser
			36520693	G => T, Thr => Lys
g27916	Scaffold-1907	chitinase-like protein 2 (fungal degradation)	79470143	T => C, Leu => Ser
g30250	Scaffold-1907	chitinase-like protein 3 (fungal degradation)	122190721	A => C, Ile => Leu
			122190722	T => A, Ile => Asn
g7251	Scaffold-1910	spätzle 1B (toll pathway)	13117496	G => A, Met => Ile
g12495	Scaffold-286	cactin (toll pathway)	43408950	A => G, Tyr => His
g9852	Scaffold-286	Heat shock protein 90 (general stress response)	2126187	T => C, Ser => Gly
			2126208	T => G, Lys => Gln
			2132993	A => T, Val => Asp
g7332	Scaffold-1910	janus kinase/hopscotch (jak/stat)	14180999	T => C, Ser => Gly
g19375	Scaffold-10	thiolester containing protein III (mark for phagocytosis)	26397814	A => G, Asn => Asp
g4321	Scaffold-1908	glutathione-S-transferse Gst (detoxification)	1278072	A => T, Leu => Phe
g11693	Scaffold-286	spätzle 2 (toll pathway)	29464475	A => G, Ile => Val
			29474935	G => A, Ser => Asn
g4723	Scaffold-1908	glutathione-S-transferse GstD6 (detoxification)	8184742	A => G, Gln => Arg
g28588	Scaffold-1907	cactin (toll pathway)	92057353	C => CCATG, Gly => fs
g11693	Scaffold-286	spätzle 2 (toll pathway)	29474979	TTGAAATAA => T, Ala => fs

A similar pattern of genotype-specific variants emerges when studying variants between *Buchnera strains* within each genotype. Genotype 1 strains separate clearly from *Buchnera* strains in other Potato aphid genotypes, whereas non-genotype 1 lines contain 692 shared variants, with 161 genes potentially affected (figure 2). This high number of non-synonymous mutation presents a large source of variation with potential effects on *Buchnera* phenotype and the aphid host, considering that the missense variants affect approximately one-third of all predicted genes in the *Buchnera* genome. One interesting variant of note is an A => C to transversion observed in all susceptible clonal lines on a gene encoding the riboflavin kinase protein *ribF* (gene ID IKDOEOGJ_00152; pos. 160,809). Differences in riboflavin concentrations are emphasised in section, and may lead greater capacity for riboflavin production in genotype 1 compared with the other susceptible genotypes.



Host aphid genotype	*Total variants*	*Missense variants*
7	715	210
2	708	209
p7	714	212
2	729	210
2	726	210
6	718	211
6	711	211
2	715	210
2	704	210
2	711	212
3	705	210
3	705	210
1	1	1
1*	0	0
1	0	0
1	2	1
1	10	8
1	5	5

Figure 2. Cladogram of *B. aphidicola* strains from different clonal lines of *M. euphorbiae*. *Buchnera* strains a degree of separation linked to host genotype, particularly in genotype 1 compared with other aphid genotypes. Within non-genotype 1 clones, genotypes 6 and 3 show distinct clusters while genotype 2 strains show within-genotype clusters, indicating the presence of sub-genotype lineages. Colours indicate host genotype. '*' denotes genotype of the reference *B. aphidicola*. Cladogram was generated using 762 variants, and rooted on the genotype 1 cluster.

5.3.3 dN/dS highlight immune function genes of interest

Only two (out of 1,792) orthologous proteins (g15621 and ACYPI56327; g14022 and ACYPI26180) between *M. euphorbiae* and *A. pisum* showed positive selection (*dS/dN* > 1) (figure 3); the functions of these two proteins could not be identified through InterProScan searches or via the Pea aphid orthologue. Core orthologues are often essential genes, and therefore usually do not accumulate missense mutations. Paralogous proteins in *M. euphorbiae*, on the other hand, show an increase in non-synonymous mutation accumulation (figure 4). Out of 1,320 paralogous groups, 44 were considered under selection through analysis of paralogue pairs (figure 5). Two of these belong to paralogue gene families with immune functionality (section 4.3.6), where one group encodes glutathione-S-transferases (g4723 and g4725) and the other group is linked to the toll pathway (g38440 and g38449). Specifically, these toll pathway genes encode dorsal, an important cell signalling molecule used for induction of multiple immune response pathways such as expression of antimicrobial peptides, lysosymes and the pro-phenoloxidase cascade (Belvin & Anderson, 1996; Gerardo *et al.,* 2010).

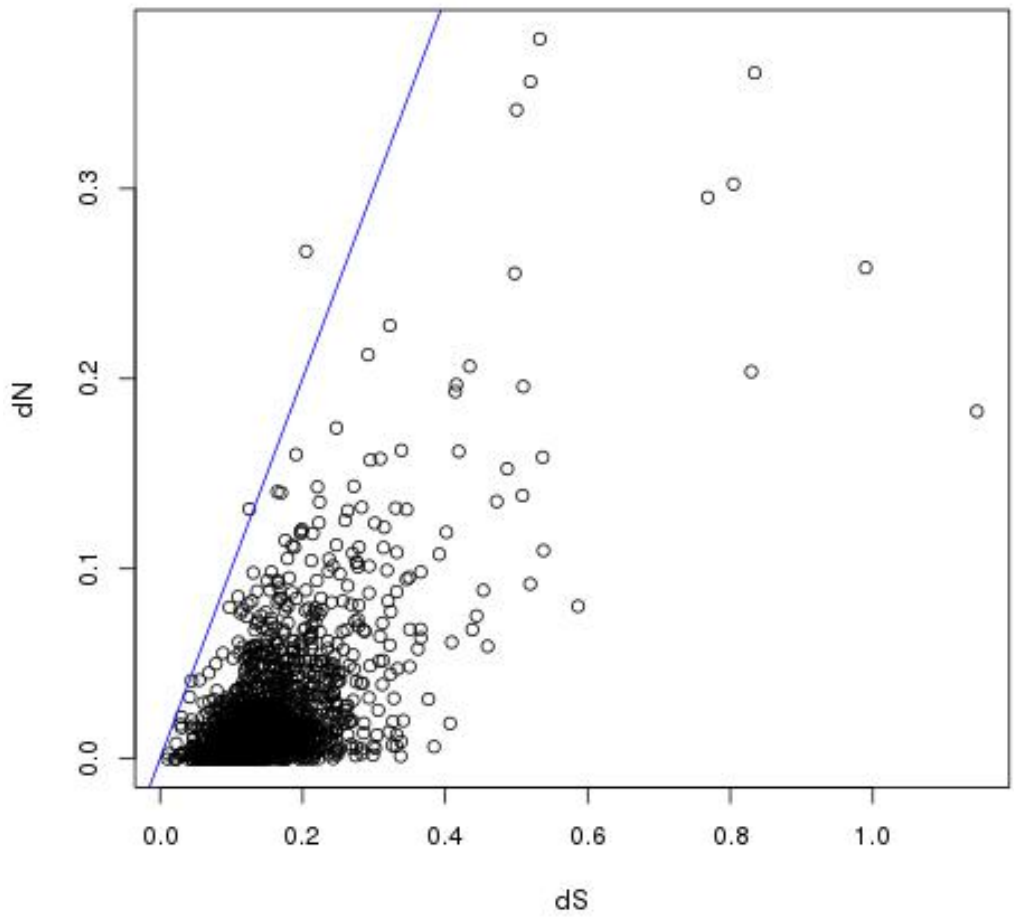


Figure 3. *dS* vs *dN* for single copy orthologues between *M. euphorbiae* and *A. pisum*. The solid blue line is *dN/dS* = 1. Data points to the left of the blue line indicate positive selection. Plot was generated from 1,792 orthologues for which *dN/dS* could be calculated. Single copy orthologues are less prone to change, gaining fewer non-synonymous changes over time.

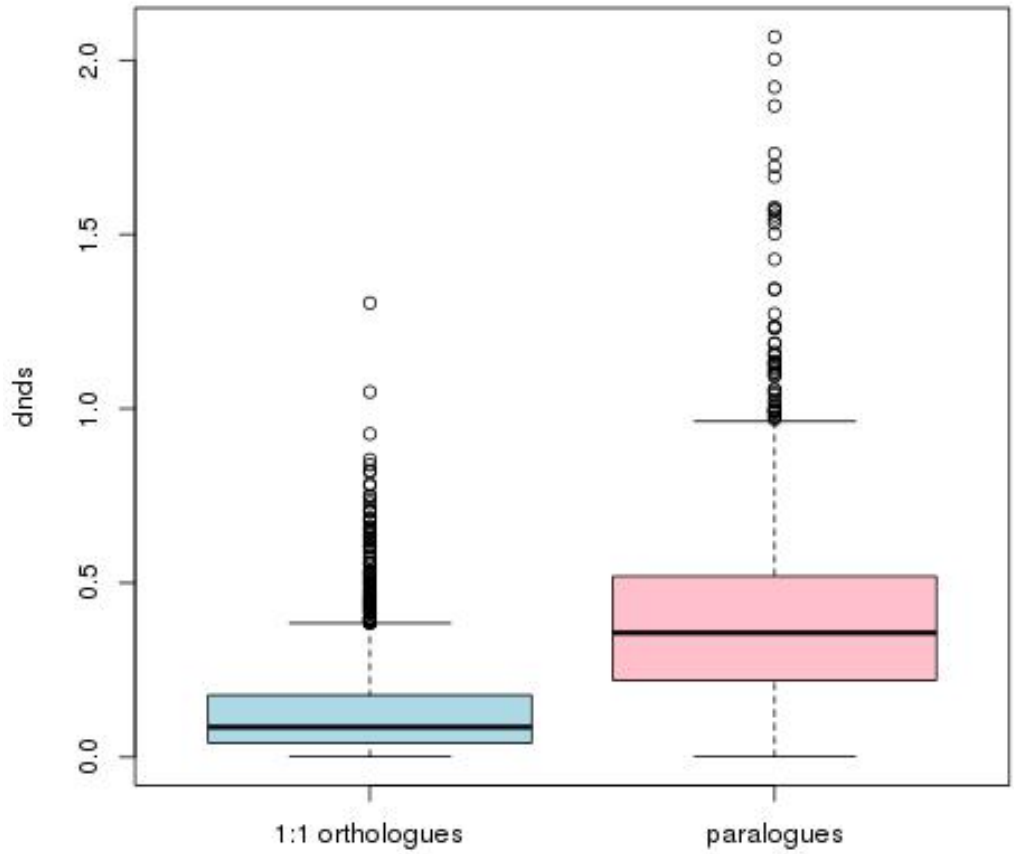


Figure 4. *dN/dS* values between *M. euphorbiae* paralogues and single copy orthologues between *M. euphorbiae* and *A. pisum* (1:1 orthologues). Shared orthologues show lower accumulation of non-synonymous mutations compared to paralogues within *M. euphorbiae*.

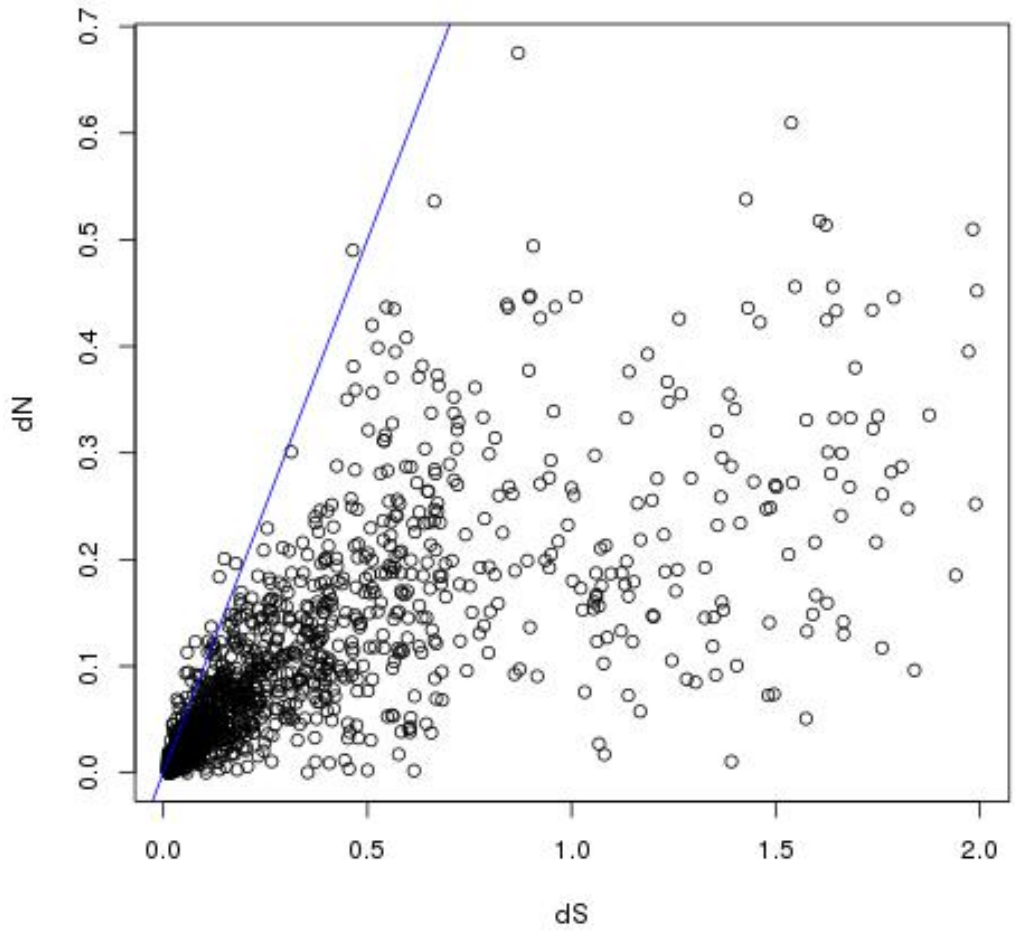


Figure 5. *dS* vs *dN* for paralogues of *M. euphorbiae*. The solid blue line is *dN/dS* = 1. Data points to the left of the blue line indicate positive selection. Plot was generated from 1,320 paralogues for which *dN/dS* could be calculated. In comparison to orthologous proteins, an increase in those with *dN/dS* > 1 is observed.

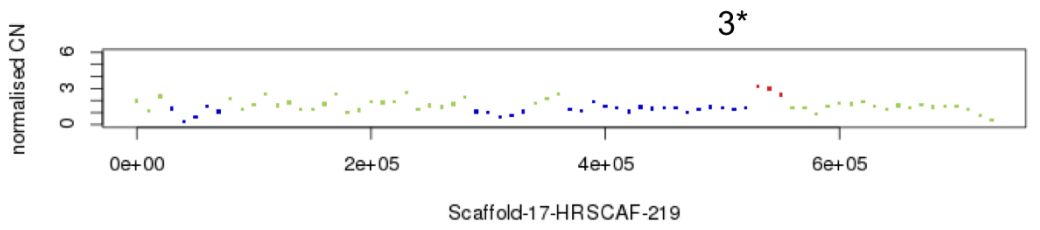
5.3.4 CNV analysis highlights increases in copy number prominent genotype 1

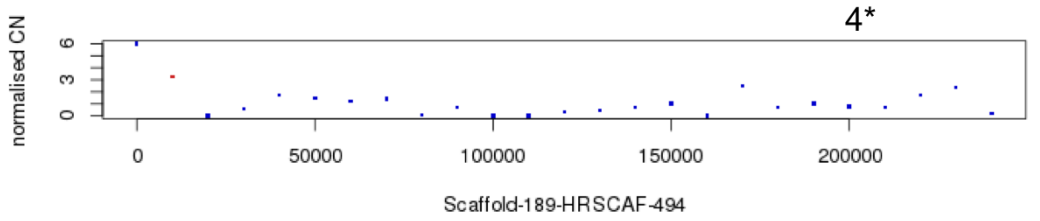
Over the genome of genotype 1 *M. euphorbiae*, 585 CNVs were detected in comparison to a pool of susceptible genotypes through control-FREEC analysis. CNV analysis identified gains or losses of 3,240 predicted genes between parasitoid-susceptible and resistant Potato aphid clones. Seven of these genes are immunity linked, having shared homology to immune related Pea aphid orthologues (table 4) (figure 6), four of which agree with gene duplications or losses observed in Orthofinder analysis against other aphid species (section 4.3.6). g15921 is a copy of proPO; susceptible genotypes possess two copies, which agrees with proPO copy number observed in other aphid species. Copy number is predicted to be lower in genotype 1, and supports Orthofinder analysis where only a single copy was identified. g36092 encodes a CLIP-domain serine protease, which induces phenoloxidase and melanisation pathway cascades (Ma *et al.,* 2019; Xu *et al.,* 2019).

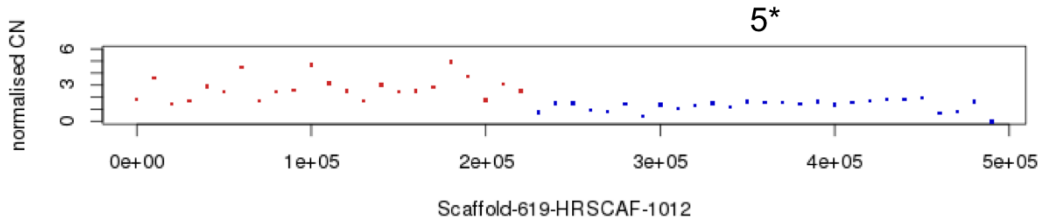
g31161, which is predicted to encode *basket*. *basket* is part of the JNK immune pathway, involved in melanisation response and wound healing (as observed in *D. melanogaster*) (Bosch *et al.,* 2005; Castillo *et al.,* 2013). *basket* has potentially been lost in susceptible genotypes (figure 6), as has been seen in *D. noxia* and duplicated in *M. persicae* (section 4.3.6). g16435 is a copy of heat shock protein 83 (Hsp83), which shows increased expression in response to a fungal pathogen in the bee species, *Megachile rotundata* (Xu & James, 2012). Finally, parasitoid resistant genotypes also appear to harbour more copies of a cytoplasmic glutathione-S-transferase GstD6 compared to susceptible clones (g4723 and g4725) (table 4).

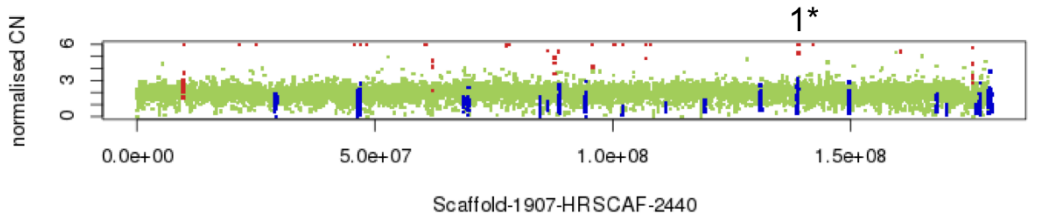
Table 4. Predicted immune genes of *M. euphorbiae* that overlap with CNVs less often found in susceptible genotypes.

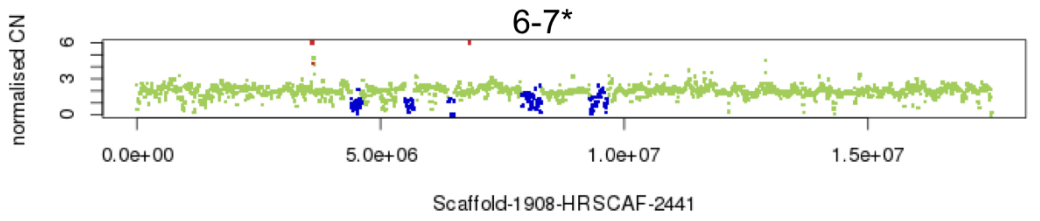
gene name	*scaffold*	*gene position*	*A. pisum homologues/predicted function*	*gain/loss in genotype 1*	*pathway*	*figure key*
g31161	Scaffold-1907	138938398-138940088	ACYPI004372 - basket	gain	jnk pathway	1
g15921	Scaffold-1918	4524813-45254260	ACYPI001367, ACYPI004484 - proPO 1/2	loss	melanisation	2
g16435	Scaffold-17	491183-493423	ACYPI002010, ACYPI009380 - hsp83	gain	stress response	3
g36092	Scaffold-189	203357-204931	ACYPI000297, ACYPI005625 - CLIP domain serine protease	gain	melanisation induction	4
g2430	Scaffold-619	356924-360039	ACYPI002386 - Jun-related antigen	gain	jnk pathway	5
g4723	Scaffold-1908	8183988-8188883	ACYPI006598 - GstD6 (glutathione-s-transferase)	gain	detoxification	6
g4725	Scaffold-1908	8206503-8211722	ACYPI006598 - GstD6 (glutathione-s-transferase)	gain	detoxification	7


3*
normalised CN
0
3
6
0e+00
2e+05
4e+05
6e+05
Scaffold-17-HRSCAF-219


4*
normalised CN
0
3
6
0
50000
100000
150000
200000
Scaffold-189-HRSCAF-494


5*
normalised CN
0
3
6
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
Scaffold-619-HRSCAF-1012


1*
normalised CN
0
3
6
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Scaffold-1907-HRSCAF-2440


6-7*
normalised CN
0
3
6
0.0e+00
5.0e+06
1.0e+07
1.5e+07
Scaffold-1908-HRSCAF-2441

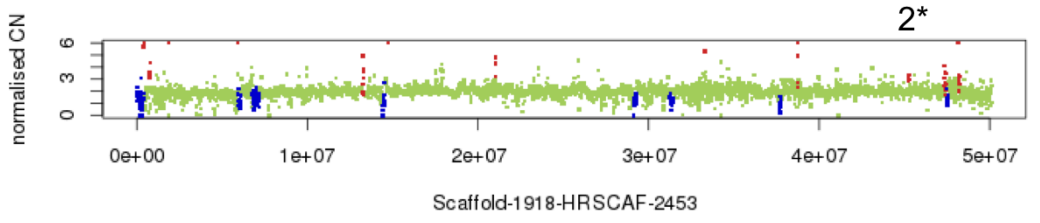

2*
normalised CN
0
3
6
0e+00
1e+07
2e+07
3e+07
4e+07
5e+07
Scaffold-1918-HRSCAF-2453

Figure 6. CNVs prominent in genotype 1 plotted against susceptible Potato aphid genotypes. Data points represent copy number (CN) over each 10,000 bp region, where green is a normal (diploid) CN of 2, blue shows reduced CN in susceptible genotypes and red shows an increased CN in susceptible genotypes. '*' see figure key column in table 4, indicating position of gene linked to the CNV.

5.3.5 Identified metabolites of M. euphorbiae

Manual curation of the aphid metabolome defined 208 spectral bins, of which 127 were annotated and provided information about 56 individual metabolites (table 5). Out of the 20 amino acids, 17 have been identified (not including aspartic acid, cysteine and serine). Unfortunately, compounds related to phenoloxidase pathways (e.g. L-DOPA and derivatives; Bakalov *et al.,* 2016; Sugumaran & Barek, 2016) were not identified. This may make it difficult to infer the role of melanization and encapsulation during parasitoid attack. However, a number of compounds linked to aphid detoxification pathways have been identified, such as glutathione, glutamate and 5-oxoproline/pyroglutamic acid.

Table 5. List of 56 metabolites in aphid tissues identified by ^{1}H-NMR.

1-Methylhistidine	L-Glutamine
2-Hydroxyvaleric acid	L-Histidine
Acetic acid	L-Isoleucine
Acetoacetic acid	L-Lactic acid
Acetylglycine	L-Leucine
Adenosine monophosphate	L-Lysine
Adenosine triphosphate	L-Methionine
Choline	L-Phenylalanine
Citric acid	L-Proline
D-Glucose	L-Threonine
D-Ribose	L-Tryptophan
D-Tartaric acid	L-Tyrosine
Dimethyl sulfone	L-Valine
Dimethylamine	N-Acetyl-L-tyrosine
Formic acid	N-nitrosdimethylamine
Fumaric acid	N2-dimethylglycine
Galactaric acid	N6-Acetyl-L-lysine
Glucaric acid	NAD
Gluconic acid	Oxypurinol
Glutathione	Phosphorylcholine
Glycerophosphocholine	Pyroglutamic acid
Glycine	Riboflavin
Imidazole	Sarcosine
Inosine	Succinic acid
L-Alanine	Trehalose
L-Arginine	Uridine
L-Asparagine	Uridine diphosphate glucuronic acid
L-Glutamic acid	Uridine diphosphate-N-acetylglucosamine

5.3.6 Variance between metabolomes of genotypes 1 and 2

Metabolomes of four-day old aphid nymphs were generated to identify any genotype specific differences in metabolite profiles. PCA revealed the first two principal components explained 49.6% of variation in aphid metabolomes. PCA analysis demonstrates a large amount of overlap between samples, with limited separation between genotypes (figure 7). Pairwise *t-test* identifies 59 peaks that differ significantly between the two genotypes (full *t-test* results are deposited at http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/volcano_table_6_jul_0hrs.csv). Plotting against log2FC (figure 8) demonstrates some of the metabolites that may be differentially elevated in genotypes 1 and 2. Genotype 1 potentially has increased levels of riboflavin (riboflavin_150; $t = 7.01$, $P < 0.01$). Riboflavin is provided to aphids via the *Buchnera* symbiont and is known to be important for nymph development (Nakabachi & Ishikawa, 1999). Genotype 1 also shows increased levels of pyroglutamate, an important glutamate store, and is derived from glutathione (Bodnaryk & McGirr, 1973; Meister, 1974), as well as a choline-like compound (Choline-like_138, $t = 7.35$, $P < 0.01$), which is linked to neuro-transmission (Martin & Krantz, 2014). Genotype 2, on the other hand, shows potentially significant increases in imidazole (imidazole_14, $t = -7.81$, $P < 0.01$) and inosine (inosine_57, $t = -5.96$, $P < 0.01$). Their roles in insects is somewhat limited, but imidazole-like compounds have previously been seen to be inhibitory to growth and development of *Bombyx mori* (Yamashita *et al.,* 1987), while inosine is linked to muscle usage in insects during periods of anoxia (depletion of oxygen) and post-anoxia recovery (Weyel & Wegenr, 1996). However, whether any of these genotype specific metabolites have any effect on aphid fitness is unclear, and would require further study.

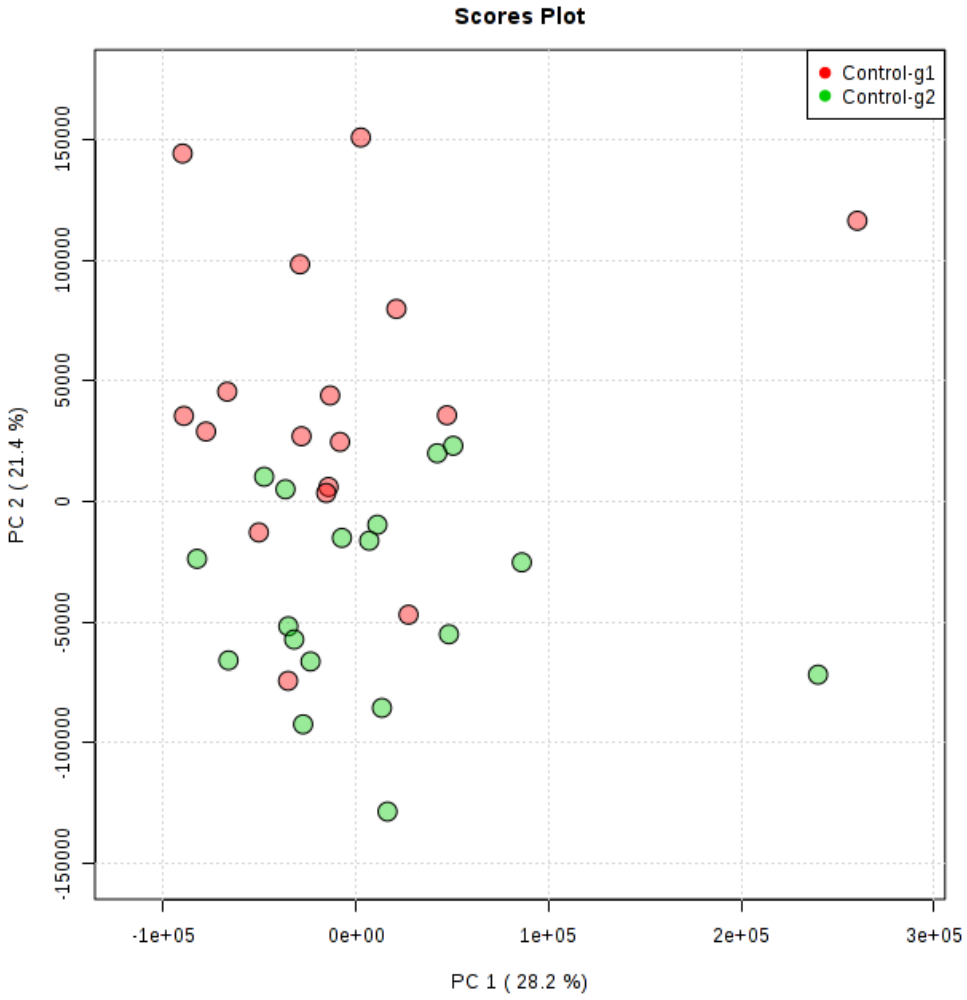


Figure 7. PCA of metabolomes of 4 day-old aphid nymphs showing scores for individuals of 'g1' and 'g2' (genotypes 1 and 2, respectively). The plot shows overlap between metabolomes of genotype 1 and 2 aphid nymphs, indicating limited differences between them.

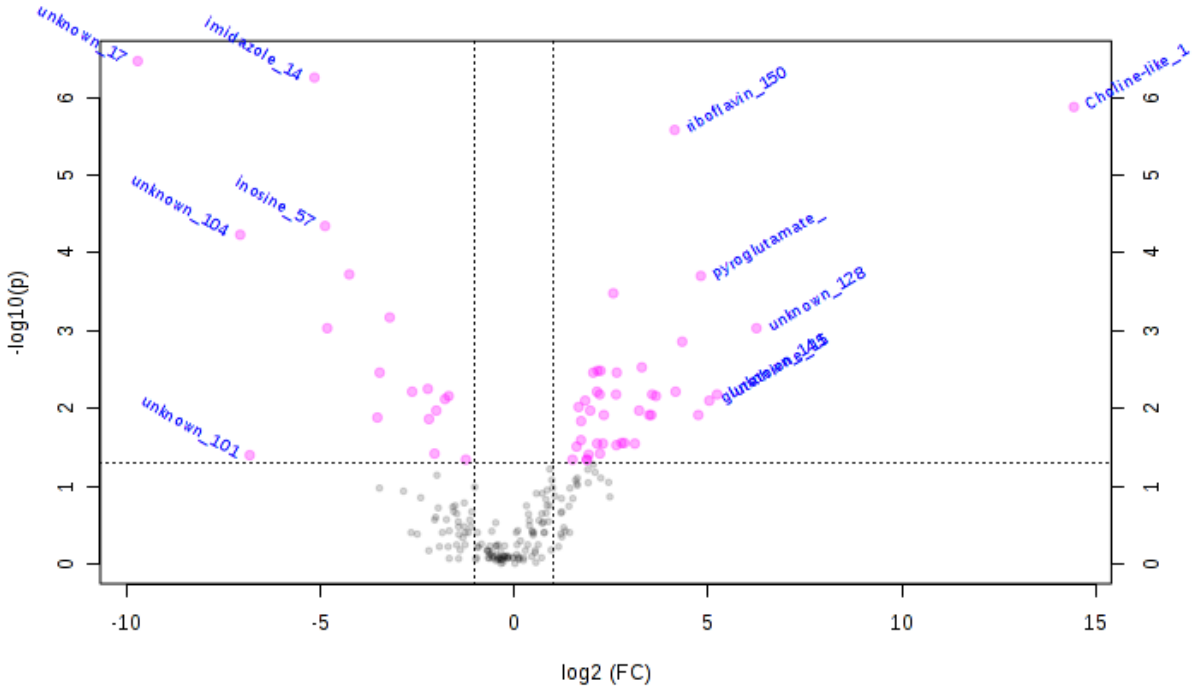


Figure 8. Volcano plot of metabolite peaks that differ significantly between aphid genotypes 1 and 2. Plotting *P* values against log2FC demonstrates further differences between metabolomes in genotype 1 and 2 nymphs.

5.3.7 Metabolomic profiles of M. euphorbiae challenged with A. ervi

A metabolomic approach was taken to assess potential up/down regulation of multiple pathways associated with parasitism challenge by *A. ervi*, by characterising the metabolomes of Potato aphid nymphs 12 hours after attack. PCA generated from metabolite peaks show that PC1, PC2 and PC3 explain 59.2% of data variation (figure 9). Overlap between samples in the four treatment categories in the score plot suggests limited differences observed between the four groups, although challenged genotypes 1 and 2 showed greatest separation compared to aphids that were unchallenged. The separation could therefore indicate that *A. ervi* challenge does have a measurable effect on aphid metabolomes.

Fluctuations in metabolites was further assessed using two-way ANOVA, from which 93 significantly different metabolite peaks were identified (figure 10) (full two-way ANOVA results are deposited at http://cgr.liv.ac.uk/pbio/MWhiteheadThesisEresources_b7dd71ba30a03ab3/two_way_anova_out.csv). The majority (50.5 %) of significantly different metabolites are a result of differences in genotype (n = 47), with far fewer linked specifically to *A. ervi* challenge (n = 6). All KEGG compounds linked to annotated metabolite peaks identified as significant (n = 67) were uploaded to KEGG mapper to identify over-represented pathways of interest. Potentially four compounds (tyrosine, fumarate, acetoacetate and succinate) are found within the tyrosine metabolism pathway (see KEGG api00350), linked to melanin synthesis and the melanisation pathway. However, all of these are significant only at the level of genotype, with genotype 1 showing increased levels of tyrosine and possibly pyroglutamate (tyrosine_49: $F = 7.29$, $P < 0.05$; pyroglutamate/succinate_161: $F = 11.07$, $P < 0.05$), while genotype 2 shows increases in levels of fumarate and acetoacetate (fumarate_54: $F = 8.96$, $P < 0.05$; acetoacetate_167: $F = 12.70$, $P < 0.01$). Also, without annotated L-DOPA metabolites and their derivatives, it is very difficult to study melanisation pathways in this fashion. Three significant metabolites (glutamate, pyroglutamate (also known as 5-oxoproline) and glutathione) are found within the glutathione metabolism pathway (see KEGG api00480), linked to the oxidative stress response and detoxification. Increased levels of two pyroglutamate peaks are found in response to *A. ervi* attack in genotype 1 specifically (two-way ANOVA pyroglutamate_160 and pyroglutamate_162, see table 6). Metabolite peaks for glutamate are not significant at the level of interaction but are significantly increased in genotype 1 at the level of treatment only (glutamate_163: $F = 29.49$, $P < 0.05$; glutamate/unknown_164: $F = 12.87$, $P < 0.05$). Finally, metabolite peaks for glutathione are only significantly increased in genotype 1 at the level of genotype (glutathione_154: $F = 14.38$, $P < 0.05$; glutathtione_152: $F = 6.36$, $P < 0.05$).

Metabolites significant across each factor assessed (n = 9) (see '*' table 6) suggest they are much more relevant to *A. ervi* resistance in genotype 1 *M. euphorbiae*, demonstrating change both in response to attack and between resistant and susceptible genotypes (sup. figure 1). Some compounds have no obvious role, especially in terms of immunity. Sarcosine (sarcosine_139) is little studied but has links to symbiont based mechanisms, where it accumulates in aposymbiotic weevils (Gasnier-Fauchet *et al.,* 1986); glutamine (glutamine_173) is linked to regulation of *B. aphidicola* titre within the aphid (Price *et al.,* 2014); while a role dimethylamine (dimetahylamine_142) is unclear. Regarding links to immunological response, decreases in pyroglutamate (pyroglutamate_160, pyroglutamate_162) are observed in challenged genotype 2, previously described as a product of glutathione metabolism and possible ROS response. Genotype 1 demonstrates increased levels of proline (proline_177), where proline has previously been described as a main constituent in antimicrobial peptides (AMPs) (Levashina *et al.,* 1995). Finally, genotype 1 challenged aphids potentially show stark increases in glucose levels (glucose_111, glucose_115), while four metabolite peaks for trehalose (trehalose_66, trehalose_88, trehalose_96 and trehalose_107) all show significant decreases in genotype 1 at the level of interaction between genotype and treatment (table 6) (sup. figure 2). Trehalose can act as energy stores for the aphid, with trehalase catabolising trehalose into glucose for use in physiological reactions (Shukla *et al.,* 2014). The specific immunological reaction this could be providing energy for though is, however, unclear.

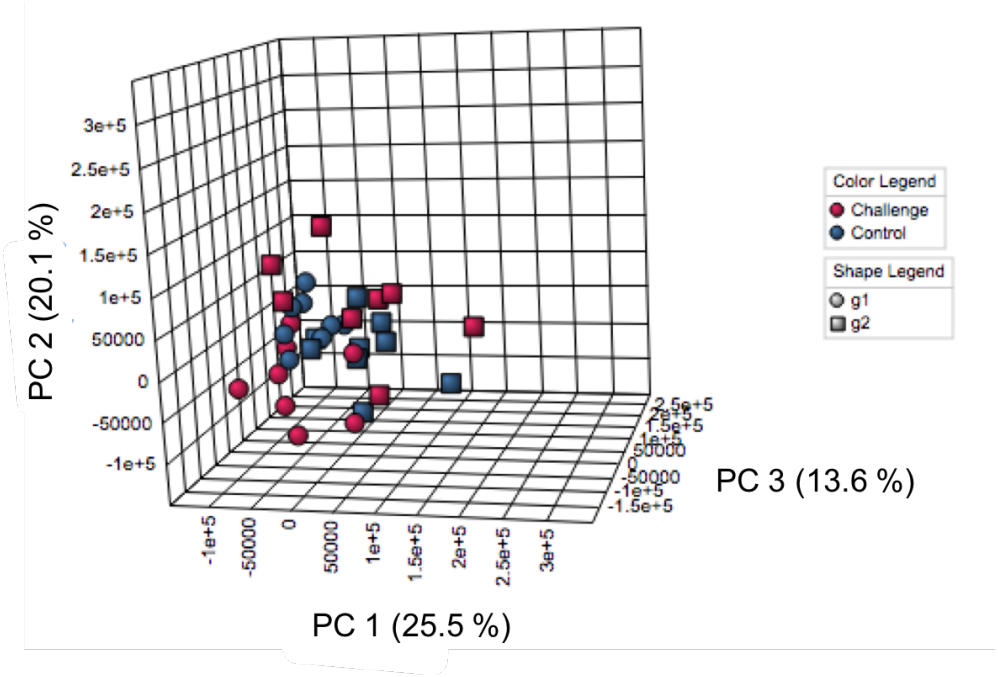


Figure 9. 3 component PCA of metabalomes showing scores for aphids 12 hours after attack by *A. ervi* (challenge) or not (control) and belonging to either 'g1' or 'g2' (genotypes 1 and 2, respectively). Challenged aphids for both *M. euphorbiae* genotypes show greater separation between each other compared to control counterparts.

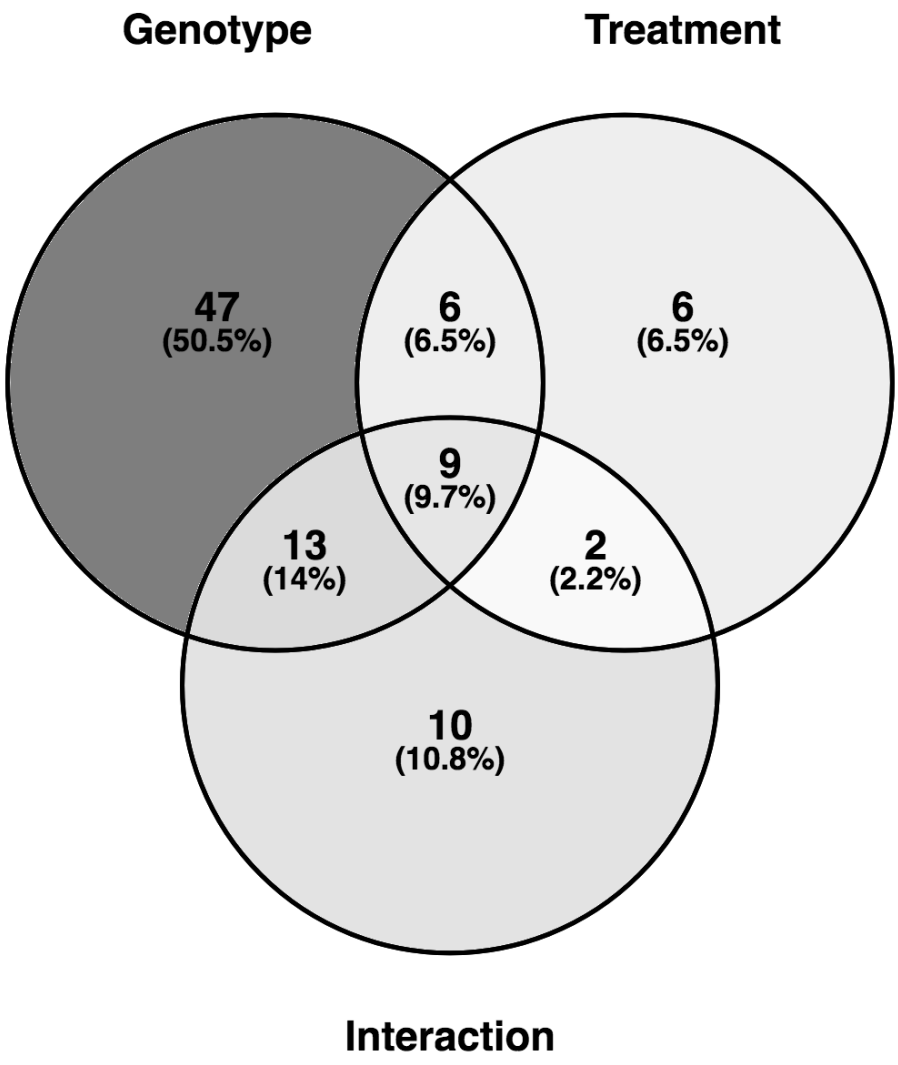


Figure 10. Venn diagram of significant metabolite peaks identified between Potato aphid genotypes 1 and 2 ("Genotype"), between *A. ervi* challenged and un-challenged aphids ("Treatment") and the interaction of genotype and treatment ("Interaction").

Table 6. Two-way ANOVA results of significantly altered metabolite peaks at level of interaction between genotype (genotype 1 and 2) and treatment (*A. ervi* challenge or no challenge). '*' denotes metabolite peaks significant at every factor (n = 9). Significance is based on false discovary rate adjusted *P* values.

metabolite peak	*genotype F value*	*genotype P value*	*genotype FDR adjusted P value*	*treatment F value*	*treatment P value*	*treatment FDR adjusted P value*	*interaction F value*	*interaction P value*	*interaction FDR adjusted P value*
O-phospho-choline_99	3.7139	0.063477	0.13898	2.1474	0.15321	0.55909	10.575	0.002832	0.023563
O-phosphocholine_117	29.047	7.75E-06	1.33E-04	4.2294	4.85E-02	0.24893	26.896	1.38E-05	0.0018619
arginine_112	0.01686	0.89755	0.91069	18.384	0.0001723	0.0039819	16.507	0.00032117	0.006073
arginine_114	0.44413	0.51023	0.5896	12.596	0.0012962	0.012255	11.455	0.0020031	0.018115
asparagine_135	8.3498	0.0071005	0.025911	2.8424	0.10218	0.43375	24.293	2.85E-05	0.0018619
* dimethylamine_142	33.239	2.69E-06	7.00E-05	25.009	2.32E-05	0.00085531	14.949	0.00055059	0.0067366
* glucose_111	16.567	0.00031469	0.0026182	17.504	0.00022993	0.0043477	8.9815	5.43E-03	0.036286
* glucose_115	9.1877	0.0049832	0.019933	13.19	0.0010386	0.012001	9.8317	0.0038214	0.027409
glutamine_159	30.198	5.76E-06	0.00011555	3.6256	0.066528	0.3145	11.577	0.0019108	0.018066
glutamine_171	7.5196	1.02E-02	3.25E-02	4.2384	4.83E-02	0.24893	8.5937	0.0063985	0.039144
* glutamine_173	10.218	0.0032673	0.014158	9.1336	5.10E-03	0.046095	10.378	0.0030634	0.024507
leucine_184	2.271	1.42E-01	2.27E-01	0.84027	0.36663	0.85684	11.694	1.83E-03	0.018066
lysine_129	1.4849	2.33E-01	0.32456	0.026503	8.72E-01	0.92853	21.459	6.57E-05	0.0027314
methionine/glutamine_172	14.033	0.00076349	0.0049627	3.9906	5.49E-02	0.26552	15.2	5.04E-04	0.0067366
phe/trp_36	0.31673	0.57776	0.63991	0.24698	0.62283	0.92135	15.198	0.00050442	0.0067366
phe/unknown_124	16.347	0.00033913	0.002713	5.7048	0.023413	0.13527	9.8574	0.0037816	0.027409

phenylalanine_35	0.95025	0.33745	0.42799	0.37203	0.5465	0.89811	13.961	0.00078351	0.0090539
* proline/glutamate_176	31.751	3.88E-06	8.98E-05	29.824	6.34E-06	0.00047849	18.301	0.000177	0.0052594
* proline_177	7.9789	0.0083332	0.027957	26.125	1.70E-05	0.00085531	13.072	0.0010851	0.011879
pyroglutamate/glutamine_158	10.312	0.0031457	0.013921	6.8998	0.013446	0.09943	17.096	0.00026337	0.006073
* pyroglutamate_160	51.946	5.07E-08	2.64E-06	14.601	0.00062274	0.0086353	23.5	3.58E-05	0.0018619
* pyroglutamate_162	29.964	6.11E-06	0.00011555	24.797	2.47E-05	0.00085531	15.382	0.000473	0.0067366
* sarcosine_139	12.228	0.0014894	0.0077448	21.648	6.20E-05	0.001612	25.54	2.00E-05	0.0018619
trehalose/glucose_106	1.3071	0.26196	0.35381	0.020021	0.88843	0.93803	12.204	0.001503	0.015631
trehalose/glucose_89	0.22902	0.63572	0.68159	0.39598	0.53393	0.89445	15.628	0.00043423	0.0067366
trehalose_107	8.1615	0.0076996	0.027572	1.517	0.22764	0.72386	17.043	0.00026809	0.006073
trehalose_66	6.445	0.016552	0.046524	1.5034	0.22969	0.72386	8.9164	0.0055824	0.036286
trehalose_88	2.392	0.13245	0.21926	0.54528	0.46599	0.86791	15.016	0.00053772	0.0067366
trehalose_96	8.0138	0.008208	0.027957	1.1811	0.28578	0.80327	20.016	0.00010241	0.0035504
tryp/tyr_39	6.7112	0.014649	0.043529	0.092824	0.76272	0.92135	9.1964	0.0049653	0.034426
unknown_126	16.09	0.00037024	0.0028522	0.13875	0.71215	0.92135	10.219	0.0032657	0.025158
unknown_128	21.895	5.76E-05	0.00079805	1.4766	0.23378	0.72576	10.627	0.0027735	0.023563
unknown_183	2.6874	0.11159	0.19838	1.144	0.29333	0.81351	16.61	0.00031016	0.006073
unknown_34	0.052665	0.82005	0.83613	0.1637	0.68864	0.92135	8.7051	0.0061031	0.038468

5.4 Discussion

5.4.1 Whole genome sequencing (WGS) helps elucidate Potato aphid genotypes

The present study aimed to further define variation between Potato aphid genotypes at the genetic level, followed by identifying specific variants to the *A. ervi* resistant genotype 1 and assess any phenotypic effect of these changes through studying the Potato aphid metabolome. As discussed in chapter 3, common genotypes of *M. euphorbiae* appeared distinct enough to discern based on a handful of polymorphic microsatellites, but this method was unable to accurately assess whether these genotypes have any genetic relation. Assessing genotype relationships using whole genome sequencing data, however, provides a much clearer picture. Specifically, high values of *Fst* and clustering of individuals in PCA analysis support strict clonality and the lack of sexual reproduction. Aphid genotypes also differed in their strain of *Buchnera* symbiont. Phylogenetic analysis of *B. aphidicola* demonstrates clustering based on host aphid genotype, apart from genotype 2 which appears to form several sub-clusters for *Buchnera*. The small level of divergence in genotype 2 lines as identified by PCA and the potential for differences in *Buchnera* within genotype 2 support the idea of sub-genotypes within Potato aphid genotypes that was alluded to in section 3.3.1.

5.4.2 Copy number aberrations and their link to multiple Potato aphid phenotypes

Obligate asexuality within the Potato aphid presents issues related to discovery of genetic variations behind phenotypes of interest, in this case parasitoid resistance. A lack of sexual recombination and the prevalence of only a handful of genotypes results in a high number of genic variants linked to specific genotypes. This results in a high number of false positives where multiple quantitative trait loci (QTLs) will be linked to genotype rather than phenotype. Asexuality also prevents the use of backcrossing as a tool to further identify QTLs within the genome. Therefore, the project has relied heavily on identifying shared variance between genotypes of *M. euphorbiae* as well as potential differences between

aphid species. The study has also focussed mainly on immune related genes, based on Pea aphid orthologues previously identified in Gerardo *et al.* (2010).

Orthologue analysis identified some orthogroups showing potato aphid specific expansion, specifically genes linked to Hsc70, glutathione-S-transferases (GstD6), and CLIP-domain serine proteases (serpins) (section 4.3.6). Surprisingly, there is also loss of one of the copies of proPO. Regarding CNV of these genes within other *M. euphorbiae* genotypes, the latter three all show similar gene gain/loss in genotype 1, suggesting variations in GstD6, serpins and proPO are less common in parasitoid susceptible genotypes. Hsc70 shows no copy number variation, therefore is likely to be a Potato aphid specific expansion. With Hsc70 potentially being linked to cold tolerance (Burton *et al.,* 1988; Rinehart *et al.,* 2000), this could be a cause of increased survival during winter and potentially relate to *M. euphorbiae's* switch to asexuality.

On the face of it, a loss of proPO may not be too concerning for the aphid, considering the melanisation pathway is believed to be attenuated to accommodate symbiosis with *Buchnera* and other secondary endosymbionts (Laughton *et al.,* 2011; Schmitz *et al.,* 2012). The larger expansion of serpins believed to activate the melanisation pathway confounds this issue, however, (assuming these are complete and not pseudogenic). Serpins are understood to consist of expanded gene families, having undergone duplication (Ross *et al.,* 2003; Yang *et al.,* 2017) and have previously been shown to mediate melanisation against eukaryotic invaders (Abraham *et al.*, 2005). Further biochemical assays may therefore be required to assess this.

GstD6 encodes a delta-class GST, linked to recent adaptation and insecticidal resistance (Enayati *et al.,* 2005; Ranson & Hemingway, 2005). Not only do we observe genotype 1 specific expansion here (g4723 and g4725) in the form of a possible tandem duplication, they belong to one of the only two sets of immune gene paralogues under positive selection. Knowing the role of *A. ervi* venom and the generation of a hospitable environment for the developing larvae (Digilio *et*

al., 2000; Colinet *et al.,* 2014), I speculate that GstD6 of *M. euphorbiae* has evolved to prevent this, perhaps through improved clearing of free-radicals built up through host-castration or increased capacity for removing exogenous compounds found in the wasp venom itself.

5.4.3 Minor variations between Potato aphid metabolomes

Initial analysis suggests that metabolomes between genotypes 1 and 2 contain enough subtle variances that they can be somewhat distinguished from one another. Most of these discriminating compounds are little studied in aphids, such as the increase in imidazole and inosine observed in genotype 2 nymphs at four days-old. Genotype 1, however, potentially demonstrates elevated levels of riboflavin. This could be supported by a SNP variant located in the riboflavin kinase gene, *ribF*, in the *Buchnera* endosymbiont specific to genotype 1. While riboflavin and *ribF* regulation is required for healthy aphids, especially in nymph and embryonic stages (Nakabachi & Ishikawa, 1999; Bermingham *et al.,* 2009; De Clerck *et al.,* 2015), Nakabachi & Ishikawa (1999) also determined excess riboflavin introduced through the diet is detrimental to aphid fitness. What effect increased riboflavin has in Potato aphid genotype 1 is unclear, but could be an example of a phenotypic effect on the host aphid due to genetic variation of the endosymbiont host.

While metabolomics can still be a powerful and cost-effective tool for studying variations in biological process between samples, it was not as effective here as hoped in studying parasitoid resistance. At the outset, with no indication for the root cause behind resistance, a broad-targeting experiment would be required to identify genes of interest. A similar experiment taking a transcriptomic approach would also be extremely costly, especially if the end-result still does not provide an answer. Therefore, a metabolomics approach was taken. Two specific pathways of interest (melanisation pathway and detoxification/ROS response) were not deemed to change in response to parasitism challenge. Understanding about the former of the two was, unfortunately, hampered by a lack of annotated

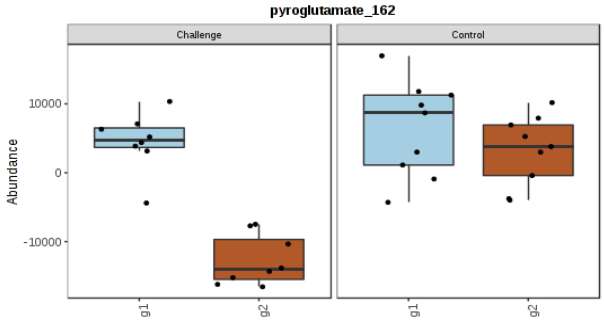
compounds such as derivatives of melanin and dopachrome that are required within the pathway.

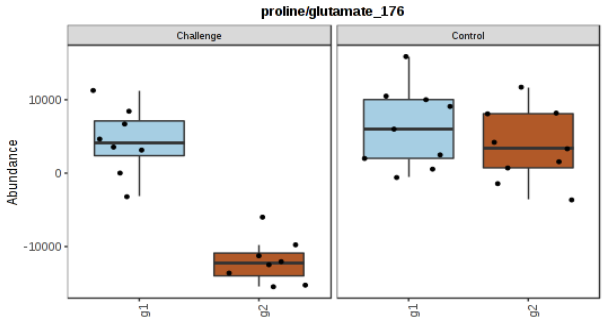
Fluctuations of sugars within the Potato aphid are observed between genotypes fllowing parasitism challenge. Specifically, genotype 1 aphids show a decrease in trehalose (a major carbohydrate energy source) as well as increasing levels of glucose in response to *A. ervi* attack. This catabolising of trehalose into glucose is similar to the response observed in *Drosophila* to fuel the cellular immune response (Bagjar *et al.,* 2015; Dolezal *et al.,* 2019). However, the attenuated immune response previously observed in aphids (Laughton *et al.,* 2011) contradicts this, and therefore suggests glucose is utilised in a different immunological response.

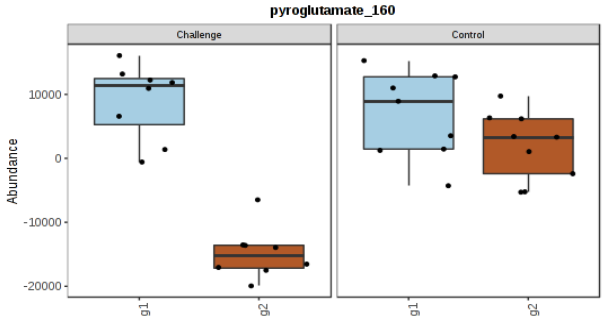
5.4.4 Conclusion

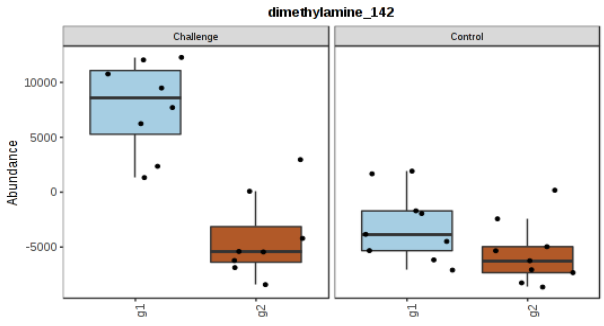
Asexuality in UK populations of *M. euphorbiae* provides challenges to identifying the genetic cause of parasitism resistance to *A. ervi*. However, through genome sequencing of the most prevalent genotypes, it is possible to imply some immune pathways of interest, such as those linked to detoxification. Metabolomics was unable to support observed genetic variance in immune genes, primarily due to difficulties in identifying metabolite compounds, but highlights some metabolome variation between genotypes 1 and 2. Going forward, theorised immune pathways potentially linked to parasitoid resistance may require the use of immunological assays, or transcriptomic studies.

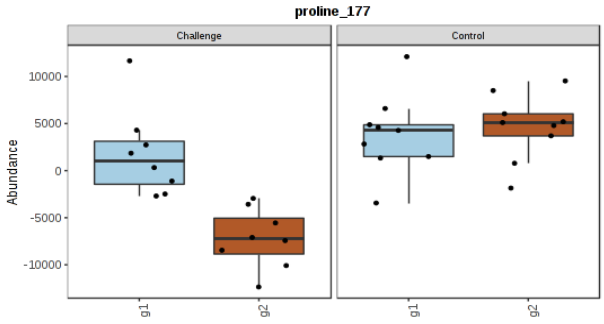
5.5 Supplementary figures

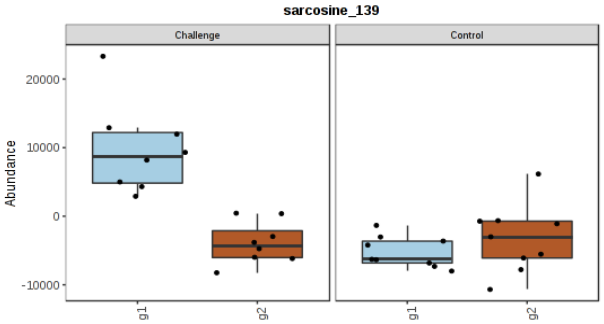


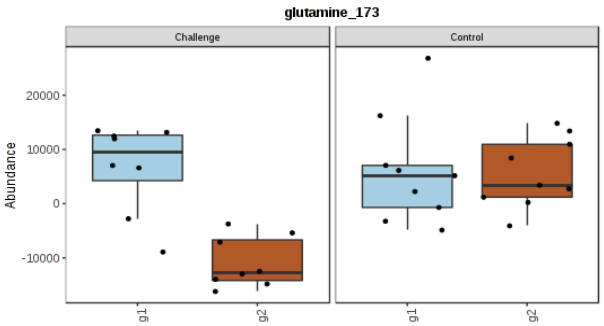


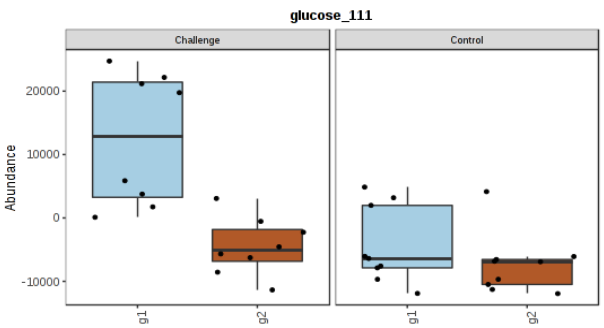


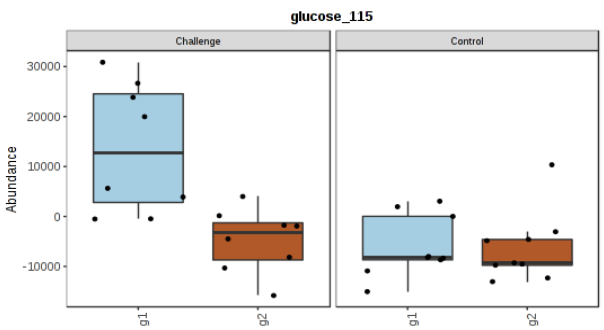




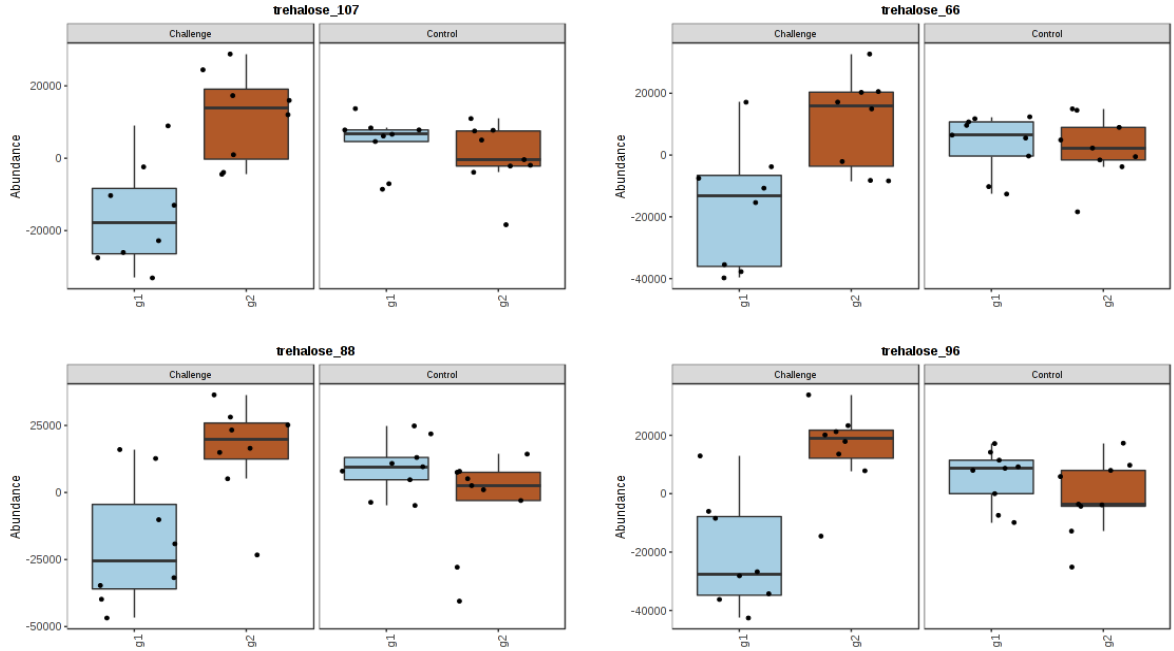








Supplementary figure 1. Select metabolite peaks showing significantly different abundance (based on metabolite peak size) across all factors tested (genotype, treatment and interaction). Genotype 1 and 2 are denoted by “g1” and “g2” respectively. “Challenge” indicates those attacked by *A. ervi* while “Control” are those not attacked. Plots generated through the MetaboAnalyst web-portal.



Supplementary figure 2. Trehalose metabolite peak abundances (based on metabolite peak size). The drop in trehalose in genotype 1 is significant at the level of interaction between *A. ervi* challenge and genotype (see table 6). Genotype 1 and 2 are denoted by “g1” and “g2” respectively. “Challenge” indicates those attacked by *A. ervi* while “Control” are those not attacked. Plots generated through the MetaboAnalyst web-portal.

6. General discussion

6.1 Summary of the work

6.1.1 Rationale for studying M. euphorbiae

Of approximately 5,000 described aphid species, 455 have the potential to cause widespread damage on crop via the spread of harmful plant pathogens (Blackman & Eastop, 2000). This study focussed specifically on one of these pests, the Potato aphid, *Macrosiphum euphorbiae,* a polyphagous aphid capable of causing reductions in crop yield through direct damage caused by feeding and as a vector or crop viruses. As observed in other aphid species, individuals within *M. euphorbiae* populations present heightened resistance to the natural enemy *Aphidius ervi* (Clarke *et al.,* 2017). Aphid resistance to parasitism in many aphid species is frequently conferred by the symbiotic relationship with aphid bacterial endosymbionts, such as *Hamiltonella defensa* for example (e.g. Oliver *et al*., 2003; Ferrari *et al.,* 2004). However, it is becoming more apparent that innate factors can also play a part (Martinez *et al.,* 2014), especially with regard to the Potato aphid (Clarke et al, 2017). Not only is it important that the genetic basis and physiological mechanisms underpinning these phenotypes are understood, it is imperative that the frequency of individuals presenting increased tolerance to control is monitored. Previous cases of aphid clones with advantageous traits, or "superclones", highlights the importance of monitoring (Chen *et al.,* 2012; Rubio-Melendez *et al.,* 2019). Ideally, identifying where pockets of tolerance against control exist can inform IPM strategies and the shift to more effective control methods. Increasing our understanding of the underlying genetic causes behind resistance to control will also drive the generation of new pest control methods that circumvent these mechanisms. For example, knowledge of the cause(s) of parasitoid resistance could be used as a basis for selective breeding for parasitoid types capable of overcoming aphid immunity.

6.1.2 Further insights into UK populations of M. euphorbiae

While *M. euphorbiae* have commonly been observed to exist as a cyclical parthenogenetic species in the USA, the majority of UK populations appear to maintain strict asexuality, as suggested by work in chapters 3 and 5. Using microsatellite analysis, apart from 16 singletons that could not be placed, Potato aphids fell into one of eight distinct genotypes. Genotype classifications were further supported by genotyping based on whole genome sequencing. Genotype 2 was the most frequently encountered genotype in native aphid populations, followed by genotype 3. Interestingly, genotype 3 was never found in a commercial setting, suggesting a role for host range in its distribution or potentially even insecticide sensitivity. Regarding frequency of *A. ervi* resistant genotype 1, its proportion within the population was surprisingly low. Where other aphids harbour such tolerances or advantageous features, it can often be associated with a fitness cost (Gwynn *et al.,* 2005; Foster *et al.,* 2010; Polin *et al.,* 2014). However, there was no evidence from this study that *A. ervi* resistant lines experienced a fitness cost in terms of post attack survival and fecundity, nor in response to exposure to sub-zero temperatures. In fact, findings suggested that adults aphids of the *A. ervi* resistant genotype have increased cold tolerance and therefore greater chance of survival during winter. These findings pose a conundrum - that the advantageous traits do not confer increased frequency of the resistant genotype in the wild and suggests another factor limits their abundance. It may be that host plant and its quality as a source of nutrition plays more of a role, as alluded to in Karley *et al.* (2017), or predatory natural enemies other than parasitoid wasps play a much larger role in curbing Potato aphid populations.

6.1.3 Contribution of the M. euphorbiae genome to the field

One of the main aspects of this project was to characterise the genetic basis of resistance to *A. ervi* in genotype 1 Potato aphids using a mixture of comparative genomics and metabolomics (chapters 4 and 5). The first step of which required

the generation of a high quality draft genome. The Potato aphid genome was constructed with a combination of "gold-standard" methods, using long-read PacBio sequence reads followed by further contig orientation and scaffolding with 10x linked reads and Hi-C scaffolding, the latter of which is becoming the standard for chromosome scale assemblies. The use of this approach allows a better understanding of genome structure and evolution between related species, especially with regards to large structural rearrangements that in turn could impact gene functionality. Currently, along with recent Hi-C integrated genomes for *Acyrthosiphon pisum* (Li *et al.,* 2019) and *Rhopalosiphum maidis* (Chen *et al.,* 2019), *M. euphorbiae* is only the third aphid genome to be assembled to such a high level.

The use of long reads and long range information was essential in assembly for the Potato aphid genome given its highly repetitive structure, with approximately 40% of the genome consisting of repeat elements (chapter 4). Characterizing repeat regions correctly could lead to identification of genes sometimes lost in short-read assemblies alone, or identify transposable elements that are often linked to increases in genome size or gene duplication/loss (Kidwell, 2002; Bast *et al.,* 2015). While not studied in-depth here, long-spanning linkage information is also useful in haplotype resolution, where single haplotype assemblies may contain a combination of different sequences from diploid chromosomal sets (or more in the case of polyploid genomes). Resolving these haplotypes correctly may uncover structural variation pertinent to the phenotypes observed in genotype 1, as previously seen in other work studying haplotype-resolved genomes (Koren *et al.,* 2018).

6.1.4 Multiple candidate causes of A. ervi resistance

The ideal approach to study the genetic basis of *A. ervi* resistance in *M. euphorbiae* would be a classical approach, performing back-crosses between susceptible and resistant lines and identifying quantitative trait loci between genomes of the parents and offspring. Unfortunately, the lack of sexual

reproduction in these lines makes this method impossible. Instead, a multi-omic approach was utilised to identify a list of candidate genes and pathways that could be implicated. Metabolomics data indicates the mounting of an immune response through the conversion of trehalose to glucose for use as an energy source, but which pathway this fuels is unclear. Increased glucose consumption is required during an encapsulation response in other insects, but the response in aphids is believed to be weak (Laughton *et al.,* 2011). Comparative genomics between other aphid species and the sequenced Potato aphid genotypes suggests a role for multiple immune pathways, such as JAK/STAT and Toll pathways, stress response pathways and genes linked to detoxification (chapters 4 and 5). The complicated nature of resistance to parasitoids (Jalvingh *et al.,* 2014; Gerritsma *et al.,* 2019) suggests it is likely that multiple pathways may be required to confer a greater chance of survival. This is also seen in other types of resistance traits, such as recently observed insecticide resistance in malaria mosquitoes requiring multiple copy number aberrations (Lucas *et al.,* 2019).

6.1.5 A possible role for Buchnera in genotypic variation

Buchnera aphidicola plays a vital role within the aphid host, providing essential nutrients the aphid it is unable to synthesise itself or obtain from its phloem sap diet. Simultaneously, the aphid is also providing the *Buchnera* with nutrition which it cannot synthesise in the form of carbon compounds and amino acids. This dependency between the aphid and the bacterial symbiont is highlighted by the tightly woven pathways of amino acid biosynthesis genes between the two organisms (Shigenobu *et al.,* 2000; IAGC, 2010). Some work, however, suggests the role of *Buchnera* could also extend beyond nutritional benefit (Dunbar *et al.,* 2007) and that *Buchnera* are specialised to their specific aphid host, even between individuals belonging to the same aphid species (Chong & Moran, 2016). At a larger evolutionary distance, it may be apparent that differences in gene repertoire of the *Buchnera* genome dictates host range amongst aphids, as hinted at by Jiang *et al.,* (2013) in relation to the role of the asparaginase *ansA* and

asparagine (a common phloem sap constituent), where *ansA* is rare in *Buchnera* genomes, but its presence may allow a more adaptive use of dietary amino acids.

Variant analysis between *Buchnera* strains belonging to different genotypes of host Potato aphid further demonstrates the clonality of these aphids, with *Buchnera* strains clustering based on the genotype of their aphid host (chapter 5). Whether genomic variants between *Buchnera* strains have differing effects on their host fitness is uncertain, but evidence from metabolomic studies of genotypes 1 and 2 indicate there are differences between variants in metabolic capacity (chapter 5). One example is the increase in riboflavin provision in genotype 1 (chapter 5), with riboflavin being an important essential nutrient in young aphid development (Nakabachi & Ishikawa, 1999).

6.2 Future work

6.2.1 Genomic and metabolomic resources for M. euphorbiae

The generation of a near-chromosome level assembly for *M. euphorbiae* is useful for the further understanding of genome evolution within aphid species, and a useful tool for scaffolding of other related organisms that may be sequenced in the future. With regard to further study of the Potato aphid genome, the reference genome generated here as well as the short-read libraries for the other most prevalent genotypes could represent most (if not all) of the current genetic composition of *M. euphorbiae* within the UK. Investigation into each genotype genome could uncover potential targets for methods of control, resulting in a more tailored approach for IPM where it is required. However, considering the dominance of genotypes 2 and 3, it may be more relevant to future control to have higher quality draft assemblies of these two populations. Both genotypes may vary in host range (Karley *et al.,* 2017; chapter 3), therefore genome assemblies may uncover other gene repertoires associated with this behaviour, such as those associated with chemosensory and olfactory proteins (Duvaux *et al.,* 2015), or aphid effector proteins (Pitino & Hogenhout, 2013).

Metabolome comparison between aphid genotypes 1 and 2 suggested a handful of compounds that varied in tissue concentrations, such as the potential difference in riboflavin provision from the *Buchnera* symbiont. The variation between *Buchnera* strains within genotypes could also be a source of variation in the aphid host, especially if ultimately it influences host development, reproductive capacity or (less likely) host plant range, and therefore may warrant further study.

6.2.2 Further study into parasitoid resistance and other resistance traits in M. euphorbiae

This project and previous work have demonstrated that genotype 1 aphids exhibit multiple advantageous fitness traits: having greater cold tolerance (chapter 3), no obvious loss of fitness after attack (chapter 3) and elevated esterase activity and therefore potentially greater tolerance to insecticides (Foster *et al.,* 2002; Clarke *et al.,* 2018), as well as *A. ervi* resistance and high reproductive capacity under benign conditions (Clarke et al., 2017). This demonstrates genotype 1 *M. euphorbiae* generally has a greater tolerance for stress from external stimuli which may require further study, especially if it is useful in IPM strategies.

With elevated tolerance to *A. ervi* attack appearing a complex trait to study, it may unfortunately require a more quantitative approach to discern the genes of interest, specifically through transcriptome studies of aphid responses following wasp challenge. The variable efficiency of successful oviposition by *A. ervi* (Sidney *et al.,* 2010), as well as the need for multiple time points and multiple aphid genotypes (both with and without attack for control purposes), would therefore require a costly and time-consuming experiment. However, the experiment may be reduced in scale should: i) the time at which an immune response occurs be pin-pointed; and ii) specific pathways of interest be identified beforehand to allow a targeted approach. As suggested in chapter 3, no apparent loss of fitness in genotype 1 would suggest that parasitism resistance results from an early response against the *A. ervi* egg and/or venom, before any negative effects such

as loss in fecundity or decreased longevity occur (Clarke *et al.,* 2017). Specific biochemical assays for phenoloxidase activity, ROS response or haemocyte counting for assessing the encapsulation response could also be used to implicate or rule out these immune pathways. Should any of these measures of immune responses be elevated in resistant lines, this may allow a targeted transcriptomic approach (e.g. via qPCR), again reducing overall time and cost.

6.2.3 Breeding programmes for overcoming resistance

As with phenotypic variation in Potato aphid genotypes, there are likely multiple phenotypes that arise within *A. ervi* as well that can overcome aphid resistance to parasitism. Perhaps understanding the parasitoid genetic changes that can lead to breaking genotype 1 resistance would provide more information on how genotype 1 Potato aphids can prevent development of the wasp larvae. However, with genotype 1 frequency being so low in the population, the requirement for *A. ervi* that can successfully parasitize this population of *M. euphorbiae* is low. While speculative, it would also remain to be seen how this may change the efficiency of *A. ervi* parasitizing other aphid genotypes, perhaps even other aphid species, the efficiencies of other parasitoid wasp species on these aphids (Alspen *et al.,* 2014), and how these could affect aphid population dynamics (Herzog *et al.,* 2007; Käch *et al.,* 2017).

6.3 General Conclusions

Despite the high level of *A. ervi* resistance in genotype 1, this does not lead to high frequency of this genotype within the tested UK Potato aphid population. This, as well as population frequencies observed in dominant genotypes 2 and 3, demonstrate the value of monitoring within-species genotypic composition and how it may inform IPM for *M. euphorbiae*. Genome and metabolome work have also highlighted how strikingly different Potato aphid genotypes can be, both at the level of the aphid and *Buchnera* symbiont. While studying this variation does not uncover the genetic basis for *A. ervi* resistance, it does provide more interesting features of *M. euphorbiae* to study going forward, as well as further investigate genome evolution in Aphididiae.

References

Abraham, E.G., Pinto, S.B., Ghosh, A., Vanlandingham, D.L., Budd, A., Higgs, S., Kafatos, F.C., Jacobs-Lorena, M. and Michel, K., 2005. An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proceedings of the National Academy of Sciences*, 102(45), pp.16327-16332.

Adams, J.B., 1962. Aphid survival at low temperatures. *Canadian Journal of Zoology*, 40(6), pp.951-956.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F. and George, R.A., 2000. The genome sequence of *Drosophila melanogaster. Science*, 287(5461), pp.2185-2195.

Alford, L., Andrade, T.O., Georges, R., Burel, F. and Van Baaren, J., 2014. Could behaviour and not physiological thermal tolerance determine winter survival of aphids in cereal fields?. *PloS one*, 9(12), p.e114982.

Altincicek, B., Gross, J. and Vilcinskas, A., 2008. Wounding-mediated gene expression and accelerated viviparous reproduction of the pea aphid *Acyrthosiphon pisum. Insect molecular biology*, 17(6), pp.711-716.

Alvarado, P., Balta, O. and Alomar, O., 1997. Efficiency of four heteroptera as predators of *Aphis gossypii* and *Macrosiphum euphorbiae* (Hom.: Aphididae). *Entomophaga*, 42(1-2), pp.215-226.

An, C., Zhang, M., Chu, Y. and Zhao, Z., 2013. Serine protease MP2 activates prophenoloxidase in the melanization immune response of *Drosophila melanogaster. PLoS One*, 8(11), p.e79533.

Andrews, M.C., Callaghan, A., Field, L.M., Williamson, M.S. and Moores, G.D., 2004. Identification of mutations conferring insecticide-insensitive AChE in the cotton-melon aphid, *Aphis gossypii* Glover. *Insect Molecular Biology*, 13(5), pp.555-561.

Asplen, M.K., Bano, N., Brady, C.M., Desneux, N., Hopper, K.R., Malouines, C., Oliver, K.M., White, J.A. and Heimpel, G.E., 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids*. Ecological Entomology*, 39(6), pp.736-739.

Bajgar, A., Kucerova, K., Jonatova, L., Tomcala, A., Schneedorferova, I., Okrouhlik, J. and Dolezal, T., 2015. Extracellular adenosine mediates a systemic metabolic switch during immune response. *PLoS biology*, 13(4), p.e1002135.

Bale, J.S., Strathdee, A.T. and Strathdee, F.C., 1994. Effects of low temperature on the arctic aphid *Acyrthosiphon brevicorne. Functional Ecology*, 8, pp.621-626.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P. and Marth G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics,* 27(12):1691-2

Barribeau, S.M., Parker, B.J. and Gerardo, N.M., 2014. Exposure to natural pathogens reveals costly aphid response to fungi but not bacteria. *Ecology and Evolution*, 4(4), pp.488-493.

Bass, C., Puinean, A.M., Zimmer, C.T., Denholm, I., Field, L.M., Foster, S.P., Gutbrod, O., Nauen, R., Slater, R. and Williamson, M.S., 2014. The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae. Insect biochemistry and molecular biology*, 51, pp.41-51.

Bast, J., Parker, D.J., Dumas, Z., Jalvingh, K.M., Tran Van, P., Jaron, K.S., Figuet, E., Brandt, A., Galtier, N. and Schwander, T., 2018. Consequences of asexuality in natural populations: insights from stick insects. *Molecular biology and evolution*, 35(7), pp.1668-1677.

Bast, J., Schaefer, I., Schwander, T., Maraun, M., Scheu, S. and Kraaijeveld, K., 2015. No accumulation of transposable elements in asexual arthropods. *Molecular biology and evolution*, 33(3), pp.697-706.

Bates, D., Mächler, M., Bolker, B. and Walker, S., 2014. Fitting linear mixed-effects models using lme4. *arXiv*:1406.5823.

Battaglia, D., Colella, T., Laurino, S., Grossi, G., Salvia, R., Riviello, L., Grimaldi, A., Congiu, T., de Eguileor, M. and Falabella, P., 2014. The effect of *Leptomastix dactylopii* parasitism and venom injection on host *Planococcus citri. Invertebrate Survival Journal*, 11(1), pp.273-2

Bayega, A., Oikonomopoulos, S., Zorbas, E., Wang, Y.C., Gregoriou, M.E., Tsoumani, K.T., Mathiopoulos, K.D. and Ragoussis, J., 2018. Transcriptome landscape of the developing olive fruit fly embryo delineated by Oxford Nanopore long-read RNA-Seq. *bioRxiv*, p.478172.

Becker, A., Schlöder, P., Steele, J.E. and Wegener, G., 1996. The regulation of trehalose metabolism in insects. *Experientia*, 52(5), pp.433-439.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), pp.289-300.

Bereswill, S., Hinkelmann, S., Kist, M. and Sander, A., 1999. Molecular Analysis of Riboflavin Synthesis Genes in *Bartonella henselae* and Use of the ribCGene for Differentiation of Bartonella Species by PCR. *Journal of clinical microbiology*, 37(10), pp.3159-3166.

Bermingham, J., Rabatel, A., Calevro, F., Viñuelas, J., Febvay, G., Charles, H., Douglas, A. and Wilkinson, T., 2009. Impact of host developmental age on the transcriptome of the symbiotic bacterium *Buchnera aphidicola* in the pea aphid (*Acyrthosiphon pisum*). *Appl. Environ. Microbiol*., 75(22), pp.7294-7297.

Birkle, L.M., Minto, L.B. and Douglas, A.E., 2002. Relating genotype and phenotype for tryptophan synthesis in an aphid–bacterial symbiosis. *Physiological entomology*, 27(4), pp.302-306.

Biteau, B., Karpac, J., Hwangbo, D. and Jasper, H., 2011. Regulation of *Drosophila* lifespan by JNK signaling. *Experimental gerontology*, 46(5), pp.349-354.

Blackman, R.L. and Eastop, V.F., 2000*. Aphids on the world's crops: an identification and information guide (No. Ed. 2).* John Wiley & Sons Ltd, New York City.

Bodnaryk, R.P. and McGirr, L., 1973. Purification, properties and function of a unique γ-glutamyl cyclotransferase from the housefly, *Musco domestica L*. *Biochimica et Biophysica Acta (BBA)-Enzymology*, 315(2), pp.352-362.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O. and Barillot, E., 2011. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, 28(3), pp.423-425.

Bosch, M., Serras, F., Martín-Blanco, E. and Baguñà, J., 2005. JNK signaling pathway required for wound healing in regenerating *Drosophila* wing imaginal discs. *Developmental biology*, 280(1), pp.73-86.

Bourguet, D., Guillemaud, T., Chevillon, C. and Raymond, M., 2004. Fitness costs of insecticide resistance in natural breeding sites of the mosquito *Culex pipiens. Evolution*, 58(1), pp.128-135.

Braendle, C., Miura, T., Bickel, R., Shingleton, A.W., Kambhampati, S. and Stern, D.L., 2003. Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS biology*, 1(1), p.e21.

Brandt, J.W., Chevignon, G., Oliver, K.M. and Strand, M.R., 2017. Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. *Proceedings of the Royal Society B: Biological Sciences*, 284(1866), p.20171925.

Broad Institute, 2019. *Picard Toolkit*, Github, http://broadinstitute.github.io/picard/;

Bruker, 2019. *TopSpin*, https://www.bruker.com/products/mr/nmr/nmr-software/nmr-software/topspin/overview.html.

Burke, G.R. and Strand, M.R., 2012. Polydnaviruses of parasitic wasps: domestication of viruses to act as gene delivery vectors. *Insects*, 3(1), pp.91-119.

Burton, V., Mitchell, H.K., Young, P. and Petersen, N.S., 1988. Heat shock protection against cold stress of *Drosophila melanogaster. Molecular and Cellular Biology*, 8(8), pp.3550-3552.

Butts, R.A., Howling, G.G., Bone, W., Bale, J.S. and Harrington, R., 1997. Contact with the host plant enhances aphid survival at low temperatures. *Ecological Entomology*, 22(1), pp.26-31.

Cabanettes, F. and Klopp, C., 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, p.e4958.

Caillaud, M.C., Boutin, M., Braendle, C. and Simon, J.C., 2002. A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). *Heredity*, 89(5), p.346.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L., 2009. BLAST+: architecture and applications. *BMC bioinformatics*, 10(1), p.421.

Cao, G., Feng, H., Guo, F., Wu, K., Li, X., Liang, G. and Desneux, N., 2014. Quantitative analysis of fitness costs associated with the development of resistance to the Bt toxin Cry1Ac in *Helicoverpa armigera. Scientific reports*, 4, p.5629.

Cardoso-Moreira, M., Arguello, J.R., Gottipati, S., Harshman, L.G., Grenier, J.K. and Clark, A.G., 2016. Evidence for the fixation of gene duplications by positive selection in *Drosophila. Genome research*, 26(6), pp.787-798.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G. and Parkhill, J., 2005. ACT: the Artemis comparison tool. *Bioinformatics*, 21(16), pp.3422-3423.

Carver, T., Thomson, N., Bleasby, A., Berriman, M. and Parkhill, J., 2008. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25(1), pp.119-120.

Castillo, J.C., Shokal, U. and Eleftherianos, I., 2013. Immune gene transcription in *Drosophila* adult flies infected by entomopathogenic nematodes and their mutualistic bacteria. *Journal of insect physiology*, 59(2), pp.179-185.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, 17(4), pp.540-552.

Chaisson, M.J. and Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC* bioinformatics, 13(1), p.238.

Chenomx, 2019*, Chenomx*, https://www.chenomx.com/.

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z. and Jander, G., 2019. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *GigaScience*, 8(4), p.giz033.

Chen, Y., Vanlerberghe-Masutti, F., Wilson, L.J., Barchia, I., McLoon, M.O., Smith, T. and Herron, G.A., 2013. Evidence of superclones in Australian cotton aphid *Aphis gossypii* Glover (Aphididae: Hemiptera). *Pest management science*, 69(8), pp.938-948.

Chen, Y., Ye, W., Zhang, Y. and Xu, Y., 2015. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic acids research*, 43(16), pp.7762-7768.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E. and Turner, S.W., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods*, 10(6), p.563.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S. and Xia, J., 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic acids research*, 46(W1), pp.W486-W494.

Chong, R.A. and Moran, N.A., 2016. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. *Proceedings of the National Academy of Sciences*, 113(46), pp.13114-13119.

Christiansen-Weniger, P. and Hardie, J., 1997. Development of the aphid parasitoid, *Aphidius ervi*, in asexual and sexual females of the pea aphid, *Acyrthosiphon pisum*, and the blackberry-cereal aphid, *Sitobion fragariae. Entomophaga*, 42(1-2), pp.165-172.

Clarke, H.V., Cullen, D., Hubbard, S.F. and Karley, A.J., 2017. Susceptibility of *Macrosiphum euphorbiae* to the parasitoid *Aphidius ervi*: larval development depends on host aphid genotype. *Entomologia Experimentalis et Applicata*, 162(2), pp.148-158.

Clarke, H.V., Foster, S.P., Oliphant, L., Waters, E.W. and Karley, A.J., 2018. Co-occurrence of defensive traits in the potato aphid *Macrosiphum euphorbiae. Ecological entomology*, 43(4), pp.538-542.

Clarkson, C.S., Temple, H.J. and Miles, A., 2018. The genomics of insecticide resistance: insights from recent studies in African malaria vectors. *Current opinion in insect science*, 27, pp.111-115.

Clough, M.S., Bale, J.S. and Harrington, R., 1990. Differential cold hardiness in adults and nymphs of the peach-potato aphid *Myzus persicae. Annals of applied biology*, 116(1), pp.1-9.

Colinet, D., Anselme, C., Deleury, E., Mancini, D., Poulain, J., Azéma-Dossat, C., Belghazi, M., Tares, S., Pennacchio, F., Poirié, M. and Gatti, J.L., 2014. Identification of the main venom protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model *Acyrthosiphon pisum. BMC genomics*, 15(1), p.342.

Coombe, L., Zhang, J., Vandervalk, B.P., Chu, J., Jackman, S.D., Birol, I. and Warren, R.L., 2018. ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC bioinformatics*, 19(1), p.234.

Costopoulos, K., Kovacs, J.L., Kamins, A. and Gerardo, N.M., 2014. Aphid facultative symbionts reduce survival of the predatory lady beetle *Hippodamia convergens*. *BMC ecology*, 14(1), p.5.

Cox, J.E., Thummel, C.S. and Tennessen, J.M., 2017. Metabolomic studies in *Drosophila. Genetics*, 206(3), pp.1169-1185.

Cui, Y., Chen, X., Luo, H., Fan, Z., Luo, J., He, S., Yue, H., Zhang, P. and Chen, R., 2016. BioCircos. js: an interactive Circos JavaScript library for biological data visualization on web applications. *Bioinformatics*, 32(11), pp.1740-1742.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15), pp.2156-2158.

Darby, A.C. and Douglas, A.E., 2003. Elucidation of the transmission patterns of an insect-borne bacterium. *Appl. Environ. Microbiol*., 69(8), pp.4403-4407.

Darby, A.C., Tosh, C.R., Walters, K.F.A. and Douglas, A.E., 2003. The significance of a facultative bacterium to natural populations of the pea aphid *Acyrthosiphon pisum*. *Ecological Entomology*, 28(2), pp.145-150.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8), pp.1164-1165.

De Clerck, C., Fujiwara, A., Joncour, P., Léonard, S., Félix, M.L., Francis, F., Jijakli, M.H., Tsuchida, T. and Massart, S., 2015. A metagenomic approach from aphid's hemolymph sheds light on the potential roles of co-existing endosymbionts. *Microbiome*, 3(1), p.63.

Degnan, P.H. and Moran, N.A., 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl. Environ. Microbiol*., 74(21), pp.6782-6791.

Delmotte, F., Leterme, N., Bonhomme, J., Rispe, C. and Simon, J.C., 2001. Multiple routes to asexuality in an aphid species. Proceedings of the Royal Society of London. *Series B: Biological Sciences*, 268(1483), pp.2291-2299.

Delmotte, F., Leterme, N., Gauthier, J.P., Rispe, C. and Simon, J.C., 2002. Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, 11(4), pp.711-723.

Devonshire, A.L., Field, L.M., Foster, S.P., Moores, G.D., Williamson, M.S. and Blackman, R.L., 1998. The evolution of insecticide resistance in the peach–potato aphid, *Myzus persicae*. Philosophical Transactions of the Royal Society of London. *Series B: Biological Sciences*, 353(1376), p

Dietzgen, R., Mann, K. and Johnson, K., 2016. Plant virus–insect vector interactions: current and potential future research directions. *Viruses*, 8(11), p.303.

Digilio, M.C., Isidoro, N., Tremblay, E. and Pennacchio, F., 2000. Host castration by *Aphidius ervi* venom proteins. *Journal of insect Physiology*, 46(6), pp.1041-1050.

Dion, E., Polin, S.E., Simon, J.C. and Outreman, Y., 2011. Symbiont infection affects aphid defensive behaviours. *Biology letters*, 7(5), pp.743-746.

Broad Institute. *DISCOVAR: Assemble genomes, find variants*. https://www.broadinstitute.org/software/discovar/blog.

Disdero, E. and Filée, J., 2017. LoRTE: Detecting transposon-induced genomic variants using low coverage PacBio long read sequences. *Mobile DNA*, 8(1), p.5.

Dixon, A.F.G., 1973. *Biology of aphids*. Edward Arnold publishers, London.

Dixon, A.F.G., Hemptinne, J.L. and Kindlmann, P., 1997. Effectiveness of ladybirds as biological control agents: patterns and processes. *Entomophaga*, 42(1-2), pp.71-83.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), pp.15-21.

Dolezal, T., Krejcova, G., Bajgar, A., Nedbalova, P. and Strasser, P., 2019. Molecular regulations of metabolism during immune response in insects. *Insect biochemistry and molecular biology*, 109, pp.31-42.

Dolgin, E.S. and Charlesworth, B., 2006. The fate of transposable elements in asexual populations. *Genetics*, 174(2), pp.817-827.

Donnelly, M.J., Isaacs, A.T. and Weetman, D., 2016. Identification, validation, and application of molecular diagnostics for insecticide resistance in malaria vectors. *Trends in Parasitology*, 32(3), pp.197-206.

Douglas, A.E. and Prosser, W.A., 1992. Synthesis of the essential amino acid tryptophan in the pea aphid (*Acyrthosiphon pisum*) symbiosis. *Journal of Insect Physiology*, 38(8), pp.565-568.

Douglas, A.E., Bouvaine, S. and Russell, R.R., 2010. How the insect immune system interacts with an obligate symbiotic bacterium. *Proceedings of the Royal Society B: Biological Sciences,* 278(1704), pp.333-338.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P. and Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333), pp.92-95.

Dudzic, J.P., Kondo, S., Ueda, R., Bergman, C.M. and Lemaitre, B., 2015. *Drosophila* innate immunity: regional and functional specialization of prophenoloxidases. *BMC biology*, 13(1), p.81.

Dunbar, H.E., Wilson, A.C., Ferguson, N.R. and Moran, N.A., 2007. Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS biology*, 5(5), p.e96.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S. and Aiden, E.L., 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems*, 3(1), pp.95-98.

Duvaux, L., Geissmann, Q., Gharbi, K., Zhou, J.J., Ferrari, J., Smadja, C.M. and Butlin, R.K., 2014. Dynamics of copy number variation in host races of the pea aphid. *Molecular biology and evolution*, 32(1), pp.63-80.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, 4(2), pp.359-361.

Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O. and Long, M., 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, 320(5883), pp.1629-1631.

Emms, D.M. and Kelly, S., 2018. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *BioRxiv*, p.466201.

Enayati, A.A., Ranson, H. and Hemingway, J., 2005. Insect glutathione transferases and insecticide resistance. *Insect molecular biology*, 14(1), pp.3-8.

Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 14(8), pp.2611-2620.

Falabella, P., Riviello, L., Caccialupi, P., Rossodivita, T., Valente, M.T., De Stradis, M.L., Tranfaglia, A., Varricchio, P., Gigliotti, S., Graziani, F. and Malva, C., 2007. A γ-glutamyl transpeptidase of *Aphidius ervi* venom induces apoptosis in the ovaries of host aphids. *Insect Biochemistry and Molecular Biology*, 37(5), pp.453-465.

Falabella, P., Tremblay, E. and Pennacchio, F., 2000. Host regulation by the aphid parasitoid *Aphidius ervi*: the role of teratocytes. *Entomologia Experimentalis et Applicata*, 97(1), pp.1-9.

Falush, D., Stephens, M. and Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4), pp.1567-1587.

Feng, R. and Isman, M.B., 1995. Selection for resistance to azadirachtin in the green peach aphid, *Myzus persicae*. Experientia, 51(8), pp.831-833.
Ferrari, J., Darby, A.C., Daniell, T.J., Godfray, H.C.J. and Douglas, A.E., 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. *Ecological Entomology*, 29(1), pp.60-65.

Ferrari, J., Godfray, H.C.J., Faulconbridge, A.S., Prior, K. and Via, S., 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution*, 60(8), pp.1574-1584.

Ferrari, J., Müller, C.B., Kraaijeveld, A.R. and Godfray, H.C.J., 2001. Clonal variation and covariation in aphid resistance to parasitoids and a pathogen. *Evolution*, 55(9), pp.1805-1814.

Ferrari, J., West, J.A., Via, S. and Godfray, H.C.J., 2012. Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. *Evolution: International Journal of Organic Evolution*, 66(2), pp.375-390.

Ffrench-Constant, R. Daborn, P.J. and Le Goff, G., 2004. The genetics and genomics of insecticide resistance. *TRENDS in Genetics*, 20(3), pp.163-170.

Foster, S. P., B. Hackett, N. Mason, G. D. Moores, D. M. Cox, J. Campbell, and I. Denholm. Resistance to carbamate, organophosphate and pyrethroid insecticides in the potato aphid (*Macrosiphum euphorbiae*). *In The BCPC Conference: Pests and diseases, Volumes 1 and 2. Proceedings of an international conference held at the Brighton Hilton Metropole Hotel*, Brighton, UK, 18-21 November 2002, pp. 811-816.

Foster, S.P., Denholm, I., Poppy, G.M., Thompson, R. and Powell, W., 2010. Fitness trade-off in peach-potato aphids (*Myzus persicae*) between insecticide resistance and vulnerability to parasitoid attack at several spatial scales. *Bulletin of entomological research*, 101(6), pp.659-666.

Foster, S.P., Tomiczek, M., Thompson, R., Denholm, I., Poppy, G., Kraaijeveld, A.R. and Powell, W., 2007. Behavioural side-effects of insecticide resistance in aphids increase their vulnerability to parasitoid attack. *Animal Behaviour*, 74(3), pp.621-632.

Frago, E., Mala, M., Weldegergis, B.T., Yang, C., McLean, A., Godfray, H.C.J., Gols, R. and Dicke, M., 2017. Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. *Nature communications*, 8(1), p.1860.

Franklin, M.T., Ritland, C.E. and Myers, J.H., 2010. Spatial and temporal changes in genetic structure of greenhouse and field populations of cabbage looper, *Trichoplusia ni. Molecular ecology*, 19(6), pp.1122-1133.

Fuchs, S., Behrends, V., Bundy, J.G., Crisanti, A. and Nolan, T., 2014. Phenylalanine metabolism regulates reproduction and parasite melanization in the malaria mosquito. P*loS one*, 9(1), p.e84865.

Fuller, S.J., Chavigny, P., Lapchin, L. and Vanlerberghe-Masutti, F., 1999. Variation in clonal diversity in glasshouse infestations of the aphid, *Aphis gossypii* Glover in southern France. *Molecular Ecology*, 8(11), pp.1867-1877.

Fukatsu, T. and Nikoh, N., 2000. Endosymbiotic Microbiota of the Bamboo Pseudococcid *Antonina crawii* (*Insecta, Homoptera*). *Appl. Environ. Microbiol*., 66(2), pp.643-650.

Garg, S., Rautiainen, M., Novak, A.M., Garrison, E., Durbin, R. and Marschall, T., 2018. A graph-based approach to diploid genome assembly. *Bioinformatics*, 34(13), pp.i105-i114.

Garrison, E. and Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv*:1207.3907.

Gasnier-Fauchet, F. and Nardon, P., 1987. Comparison of sarcosine and methionine sulfoxide levels in symbiotic and aposymbiotic larvae of two sibling species, *Sitophilus oryzae L.* and *S. zeamais* Mots. (*Coleoptera: Curculionidae*). *Insect biochemistry*, 17(1), pp.17-20.

Gavrilov-Zimin, I.A., Stekolshchikov, A.V. and Gautam, D.C., 2015. General trends of chromosomal evolution in *Aphidococca* (*Insecta, Homoptera, Aphidinea & Coccinea*). *Comparative cytogenetics*, 9(3), p.335.

Gerardo, N.M., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S.M., De Vos, M., Duncan, E.J., Evans, J.D., Gabaldón, T., Ghanim, M. and Heddi, A., 2010. Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome biology*, 11(2), p.R21.

Gerritsma, S., Jalvingh, K.M., van de Beld, C., Beerda, J., van de Zande, L., Vrieling, K. and Wertheim, B., 2019. Natural and artificial selection for parasitoid resistance in *Drosophila melanogaster* leave different genetic signatures. *Frontiers in Genetics,* 10, p479.

Ghurye, J., Pop, M., Koren, S., Bickhart, D. and Chin, C.S., 2017. Scaffolding of long read assemblies using long range contact information. *BMC genomics*, 18(1), p.527.

Girirajan, S., Campbell, C.D. and Eichler, E.E., 2011. Human copy number variation and complex genetic disease. *Annual review of genetics*, 45, pp.203-226.

Godfray, H.C.J., 1994. *Parasitoids: behavioral and evolutionary ecology*. Princeton University Press.

Gonzalez, E.A., Garg, A., Tang, J., Nazario-Toole, A.E. and Wu, L.P., 2013. A glutamate-dependent redox system in blood cells is integral for phagocytosis in *Drosophila melanogaster. Current Biology*, 23(22), pp.2319-2324.

Gorman, M.J. and Paskewitz, S.M., 2001. Serine proteases as mediators of mosquito immune responses. *Insect biochemistry and molecular biology*, 31(3), pp.257-262.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), p.644.

Grau, J.H., Hackl, T., Koepfli, K.P. and Hofreiter, M., 2018. Improving draft genome contiguity with reference-derived in silico mate-pair libraries. *GigaScience*, 7(5), p.giy029.

Gu, H., Du, J., Neto, F.C., Carroll, P.A., Turner, S.J., Chiorean, E.G., Eisenman, R.N. and Raftery, D., 2015. Metabolomics method to comprehensively analyze amino acids in different domains. *Analyst*, 140(8), pp.2726-2734.

Guillemaud, T., Mieuzet, L. and Simon, J.C., 2003. Spatial and temporal genetic variability in French populations of the peach–potato aphid, *Myzus persicae. Heredity*, 91(2), p.143.

Gwynn, D.M., Callaghan, A., Gorham, J., Walters, K.F.A. and Fellowes, M.D.E., 2005. Resistance is costly: trade-offs between immunity, fecundity and survival in the pea aphid. *Proceedings of the Royal Society B: Biological Sciences*, 272(1574), pp.1803-1808.

Haney, R.A., Clarke, T.H., Gadgil, R., Fitzpatrick, R., Hayashi, C.Y., Ayoub, N.A. and Garb, J.E., 2016. Effects of gene duplication, positive selection, and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome biology and evolution,* 8(1), pp.228-242.

Han, G., Zhang, N., Xu, J., Jiang, H., Ji, C., Zhang, Z., Song, Q., Stanley, D., Fang, J. and Wang, J., 2019. Characterization of a novel Helitron family in insect genomes: insights into classification, evolution and horizontal transfer. *Mobile DNA*, 10(1), p.25.

Hansen, A.K., Vorburger, C. and Moran, N.A., 2012. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, 22(1), pp.106-114.

Harrington, R. and Woiwod, I., 2007. Foresight from hindsight: the Rothamsted insect survey. Outlooks on pest management, 18(1), p.9. Hansen, A.K. and Moran, N.A., 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences*, 108(7), pp.2849-2854.

Harvey, T.L. and Martin, T.J., 1988. Relative cold tolerance of Russian wheat aphid and biotype-E greenbug (*Homoptera*: *Aphididae*). *Journal of the Kansas Entomological Society*, pp.137-140.

Henry-Halldin, C.N., Nadesakumaran, K., Keven, J.B., Zimmerman, A.M., Siba, P., Mueller, I., Hetzel, M.W., Kazura, J.W., Thomsen, E., Reimer, L.J. and Zimmerman, P.A., 2012. Multiplex assay for species identification and monitoring of insecticide resistance in *Anopheles punctulatus* group populations of Papua New Guinea. *The American journal of tropical medicine and hygiene,* 86(1), pp.140-151.

Henry, L., Schwander, T. and Crespi, B.J., 2011. Deleterious mutation accumulation in asexual Timema stick insects. *Molecular biology and evolution*, 29(1), pp.401-408.

Herzog, J., Müller, C.B. and Vorburger, C., 2007. Strong parasitoid-mediated selection in experimental populations of aphids. *Biology Letters*, 3(6), pp.667-669.

Heyworth, E.R. and Ferrari, J., 2015. A facultative endosymbiont in aphids can provide diverse ecological benefits. *Journal of evolutionary biology*, 28(10), pp.1753-1760.

Hickman, J.M. and Wratten, S.D., 1996. Use of *Phelia tanacetifolia* strips to enhance biological control of aphids by overfly larvae in cereal fields. *Journal of Economic Entomology*, 89(4), pp.832-840.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. (2015). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32(5):767-769.

Holt, C. and Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, 12(1), p.491.

Hong, J.W., Hendrix, D.A., Papatsenko, D. and Levine, M.S., 2008. How the Dorsal gradient works: insights from postgenome technologies. *Proceedings of the National Academy of Sciences,* 105(51), pp.20072-20076.

Hoxmeier, J.C., Thompson, B.D., Broeckling, C.D., Small, P., Foy, B.D., Prenni, J. and Dobos, K.M., 2015. Analysis of the metabolome of *Anopheles gambiae* mosquito after exposure to *Mycobacterium ulcerans. Scientific reports*, 5, p.9242.

Huang, S., Kang, M. and Xu, A., 2017. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, 33(16), pp.2577-2579.

Hufbauer, R.A. and Via, S., 1999. Evolution of an aphid-parasitoid interaction: variation in resistance to parasitism among aphid populations specialized on different plants. *Evolution*, 53(5), pp.1435-1445.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

i5K Consortium, 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), pp.595-600.

International Aphid Genomics Consortium (IAGC), 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum. PLoS biology*, 8(2), p.e1000313.

Jalvingh, K.M., Chang, P.L., Nuzhdin, S.V. and Wertheim, B., 2014. Genomic changes under rapid evolution: selection for parasitoid resistance. *Proceedings of the Royal Society B: Biological Sciences*, 281(1779), p.20132303.

Jansen, J.P., 2005. Aphid parasitoid complex in potato in the context of IPM in Belgium. *Communications in agricultural and applied biological sciences*, 70(4), pp.539-546.

Jaquiéry, J., Stoeckel, S., Nouhaud, P., Mieuzet, L., Mahéo, F., Legeai, F., Bernard, N., Bonvoisin, A., Vitalis, R. and Simon, J.C., 2012. Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Molecular Ecology*, 21(21), pp.5251-5264.

Jaquiéry, J., Stoeckel, S., Rispe, C., Mieuzet, L., Legeai, F. and Simon, J.C., 2011. Accelerated evolution of sex chromosomes in aphids, an X0 system. *Molecular biology and evolution*, 29(2), pp.837-847.

Jaron, K.S., Bast, J., Ranallo-Benavidez, T.R., Robinson-Rechavi, M. and Schwander, T., 2018. Genomic features of asexual animals. *BioRxiv*, p.497495.

Jervis, M., Kidd, N., 1996. *Insect natural Enemies: Practical approaches to their study and evalutation*. Chapman & Hall, London, UK.

Jiang, Z., Jones, D.H., Khuri, S., Tsinoremas, N.F., Wyss, T., Jander, G. and Wilson, A.C., 2013. Comparative analysis of genome sequences from four strains of the *Buchnera aphidicola* Mp endosymbiont of the green peach aphid, *Myzus persicae. BMC genomics*, 14(1), p.917.

Johnson, R.M., Ellis, M.D., Mullin, C.A. and Frazier, M., 2010. Pesticides and honey bee toxicity–USA. *Apidologie*, 41(3), pp.312-331.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. and Pesseat, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), pp.1236-1240.

Joseph, J.R., Ameline, A. and Couty, A., 2011. Effects on the aphid parasitoid *Aphidius ervi* of an insecticide (Plenum®, pymetrozine) specific to plant-sucking insects. *Phytoparasitica*, 39(1), pp.35-41.

Joshi NA, Fass JN., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. *Github*, https://github.com/najoshi/sickle.

Käch, H., Mathé-Hubert, H., Dennis, A.B. and Vorburger, C., 2018. Rapid evolution of symbiont-mediated resistance compromises biological control of aphids by parasitoids. *Evolutionary applications*, 11(2), pp.220-230.

Käch, H., Mathé-Hubert, H., Dennis, A.B. and Vorburger, C., 2018. Rapid evolution of symbiont-mediated resistance compromises biological control of aphids by parasitoids. *Evolutionary applications*, 11(2), pp.220-230.

Kacsoh, B.Z. and Schlenke, T.A., 2012. High hemocyte load is associated with increased resistance against parasitoids in *Drosophila suzukii*, a relative of *D. melanogaster*. *PloS one*, 7(4), p.e34721.

Kaiser, M.C. and Heimpel, G.E., 2016. Parasitoid-induced transgenerational fecundity compensation in an aphid. *Entomologia Experimentalis et Applicata*, 159(2), pp.197-206.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BLASTKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology*. 428, 726-731.

Karley, A.J., Emslie-Smith, M. and Bennett, A.E., 2017. Potato aphid *Macrosiphum euphorbiae* performance is determined by aphid genotype and not mycorrhizal fungi or water availability. *Insect science*, 24(6), pp.1015-1024.

Karley, A.J., Pitchford, J.W., Douglas, A.E., Parker, W.E. and Howardh, J.J., 2003. The causes and processes of the mid-summer population crash of the potato aphids *Macrosiphum euphorbiae* and *Myzus persicae* (*Hemiptera: Aphididae*). *Bulletin of Entomological Research*, 93(5), pp.425-438.

Karlsson, E., Lärkeryd, A., Sjödin, A., Forsman, M. and Stenberg, P., 2015. Scaffolding of a bacterial genome using MinION nanopore sequencing. *Scientific reports*, 5, p.11996.

Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115(1), pp.49-63.

Kim-Jo, C., Gatti, J.L. and Poirié, M., 2019. *Drosophila* cellular immunity against parasitoid wasps: a complex and time-dependent process. *Frontiers in physiology*, 10, p.603.

Kirkpatrick, M. and Barton, N., 2006. Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), pp.419-434.

Kogan, M., 1998. Integrated pest management: historical perspectives and contemporary developments. *Annual review of entomology*, 43(1), pp.243-270.

Kondrashov, F.A., 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), pp.5048-5057.

Koren, S., Rhie, A., Walenz, B.P., Dilthey, A.T., Bickhart, D.M., Kingan, S.B., Hiendleder, S., Williams, J.L., Smith, T.P. and Phillippy, A.M., 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nature biotechnology*, 36(12), p.1174.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, 27(5), pp.722-736.

Korf, I., 2004. Gene finding in novel genomes. *BMC bioinformatics*, 5(1), p.59.

Koštál, V. and Tollarová-Borovanská, M., 2009. The 70 kDa heat shock protein assists during the repair of chilling injury in the insect, *Pyrrhocoris apterus. PloS one*, 4(2), p.e4546.

Kozlov, A., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A., 2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *bioRxiv*, p.447110.

Labbé, P., Berthomieu, A., Berticat, C., Alout, H., Raymond, M., Lenormand, T. and Weill, M., 2007. Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito Cu*lex pipiens. Molecular Biology and Evolution*, 24(4), pp.1056-1067.

Laetsch, D.R. and Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*, 6.

Lambertini, C., Riis, T., Olesen, B., Clayton, J.S., Sorrell, B.K. and Brix, H., 2010. Genetic diversity in three invasive clonal aquatic species in New Zealand. *BMC genetics*, 11(1), p.52.

Lane, D.J., 1991. 16S/23S rRNA sequencing. *In Nucleic acid techniques in bacterial systematics*. New York, NY: John Wiley and Sons.

Laughton, A.M., Garcia, J.R., Altincicek, B., Strand, M.R. and Gerardo, N.M., 2011. Characterisation of immune responses in the pea aphid, Ac*yrthosiphon pisum. Journal of insect physiology*, 57(6), pp.830-839.

Laughton, A.M., Garcia, J.R. and Gerardo, N.M., 2016. Condition-dependent alteration of cellular immunity by secondary symbionts in the pea aphid, *Acyrthosiphon pisum*. *Journal of insect physiology*, 86, pp.17-24.

Lê Cao, K.A., Boitard, S. and Besse, P., 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics*, 12(1), p.253.

Leclair, M., Polin, S., Jousseaume, T., Simon, J.C., Sugio, A., Morlière, S., Fukatsu, T., Tsuchida, T. and Outreman, Y., 2017. Consequences of coinfection with protective symbionts on the host phenotype and symbiont titres in the pea aphid system. *Insect science*, 24(5), pp.798-808.

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., McCombie, W.R. and Schatz, M., 2016. Third-generation sequencing and the future of genomics. *BioRxiv*, p.048603.

Lees, A.D., 1967. The production of the apterous and alate forms in the aphid *Megoura viciae* Buckton, with special reference to the role of crowding. *Journal of Insect Physiology*, 13(2), pp.289-318.

Legarrea, S., Diaz, B.M., Plaza, M., Barrios, L., Morales, I., Viñuela, E. and Fereres, A., 2012. Diminished UV radiation reduces the spread and population density of *Macrosiphum euphorbiae* (Thomas)[*Hemiptera: Aphididae*] in lettuce crops. *Horticultural Science*, 39(2), pp.74-80.

Legeai, F., Shigenobu, S., Gauthier, J.P., Colbourne, J., Rispe, C., Collin, O., Richards, S., Wilson, A.C., Murphy, T. and Tagu, D., 2010. AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect molecular biology*, 19, pp.5-12.

Lemaitre, B. and Hoffmann, J., 2007. The host defense of *Drosophila melanogaster. Annu. Rev. Immunol*., 25, pp.697-743.

Levashina, E.A., Ohresser, S., Bulet, P., Reichhart, J.M., Hetru, C. and Hoffmann, J.A., 1995. Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with antibacterial and antifungal properties. *European Journal of Biochemistry,* 233(2), pp.694-700.

Liao, Y., Smyth, G.K. and Shi, W., 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), pp.923-930.

Libbrecht, R., Gwynn, D.M. and Fellowes, M.D.E., 2007. *Aphidius ervi* preferentially attacks the green morph of the pea aphid, *Acyrthosiphon pisum*. *Journal of insect* behavior, 20(1), pp.25-32.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), pp.3094-3100.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), pp.1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009).

The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078-9.

Li, S., Falabella, P., Giannantonio, S., Fanti, P., Battaglia, D., Digilio, M.C., Völkl, W., Sloggett, J.J., Weisser, W. and Pennacchio, F., 2002. Pea aphid clonal resistance to the endophagous parasitoid *Aphidius ervi. Journal of Insect Physiology*, 48(10), pp.971-980.

Liu, L.J., Zheng, H.Y., Jiang, F., Guo, W. and Zhou, S.T., 2014. Comparative transcriptional analysis of asexual and sexual morphs reveals possible mechanisms in reproductive polyphenism of the cotton aphid. *PloS one*, 9(6), p.e99506.

Liu, N., 2015. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annual review of entomology*, 60, pp.537-559.

Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), pp.1658-1659.

Li, Y., Park, H., Smith, T.E. and Moran, N.A., 2019. Gene Family Evolution in the Pea Aphid Based on Chromosome-level Genome Assembly. *Molecular Biology and Evolution*. msz138.

Li, Z. and Srivastava, P., 2003. Heat-shock proteins. *Current protocols in immunology*, 58(1), pp.A-1T.

Livera, A.M.D., Sysi-Aho, M., Jacob, L., Gagnon-Bartsch, J.A., Castillo, S., Simpson, J.A. and Speed, T.P., 2015. Statistical methods for handling unwanted variation in metabolomics data. *Analytical chemistry*, 87(7), pp.3606-3615.

Longley, M., 1999. A review of pesticide effects upon immature aphid parasitoids within mummified hosts. *International Journal of Pest Management*, 45(2), pp.139-145.

Lucas, E.R., Miles, A., Harding, N.J., Clarkson, C.S., Lawniczak, M.K., Kwiatkowski, D.P., Weetman, D., Donnelly, M.J. and Anopheles gambiae 1000 Genomes Consortium, 2019. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome research*, 29(8), pp.1250-1261.

Lu, Y., Wu, K., Jiang, Y., Guo, Y. and Desneux, N., 2012. Widespread adoption of Bt cotton and insecticide decrease promotes biocontrol services. *Nature*, 487(7407), p.362.

MacGillivray, M.E. and Anderson, G.B., 1964. The effect of photoperiod and temperature on the production of gamic and agamic forms in *Macrosiphum euphorbiae* (Thomas). *Canadian Journal of Zoology*, 42(3), pp.491-510.

Ma, L., Chen, F., Wang, W., Xu, L. and Lu, Z.Q., 2019. Identification of two clip-domain serine proteases involved in the pea aphid's defense against bacterial and fungal infection. *Insect science*. 00, pp1-10.

Martin, C.A. and Krantz, D.E., 2014. *Drosophila melanogaster* as a genetic model system to study neurotransmitter transporters. *Neurochemistry international*, 73, pp.71-88.

Martinez, A.J., Ritter, S.G., Doremus, M.R., Russell, J.A. and Oliver, K.M., 2014. Aphid-encoded variability in susceptibility to a parasitoid. *BMC evolutionary biology*, 14(1), p.127.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), pp.10-12.

Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O. and Dalmay, T., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, 18(1), p.27.

Mavridis, K., Wipf, N., Müller, P., Traoré, M., Muller, G. and Vontas, J., 2018. Detection and Monitoring of Insecticide Resistance Mutations in Anopheles gambiae: Individual vs. Pooled Specimens. *Genes*, 9(10), p.479.

McCarroll, S.A. and Altshuler, D.M., 2007. Copy-number variation and association studies of human disease. *Nature genetics*, 39(7s), p.S37.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297-1303.

McLean, A.H., Parker, B.J., Hrček, J., Kavanagh, J.C., Wellham, P.A. and Godfray, H.C.J., 2018. Consequences of symbiont co-infections for insect host phenotypes. *Journal of animal ecology*, 87(2), pp.478-488.

Mehrparvar, M., Zytynska, S.E. and Weisser, W.W., 2013. Multiple cues for winged morph production in an aphid metacommunity. *PloS one*, 8(3), p.e58323.

Meister, A., 1974. Glutathione; Metabolism and function via the γ-glutamyl cycle. *Life sciences*, 15(2), pp.177-190.

Michels, G.J., Elliott, N.C., Romero, R.A., Owings, D.A. and Bible, J.B., 2001. Impact of indigenous coccinellids on Russian wheat aphids and greenbugs (*Homoptera: Aphididae*) infesting winter wheat in the Texas Panhandle. *Southwestern Entomologist*, 26(2), pp.97-114.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. and Gurevich, A., 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, 34(13), pp.i142-i150.

Miller, N.J., Birley, A.J., Overall, A.D.J. and Tatchell, G.M., 2003. Population genetic structure of the lettuce root aphid, *Pemphigus bursarius* (L.), in relation to geographic distance, gene flow and host plant usage. *Heredity*, 91(3), p.217.

Mills, N.J., 2017. Rapid evolution of resistance to parasitism in biological control. *Proceedings of the National Academy of Sciences*, 114(15), pp.3792-3794.

Miura, T., Braendle, C., Shingleton, A., Sisk, G., Kambhampati, S. and Stern, D.L., 2003. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (*Hemiptera: Aphidoidea*). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 295(1), pp.59-81.

Monti, V., Manicardi, G.C. and Mandrioli, M., 2011. Cytogenetic and molecular analysis of the holocentric chromosomes of the potato aphid *Macrosiphum euphorbiae* (Thomas, 1878). *Comparative Cytogenetics*, 5(3), p.163.

Montllor, C.B., Maxmen, A. and Purcell, A.H., 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecological Entomology*. 27, 189-195.

Moran, N.A. and Jarvik, T., 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, 328(5978), pp.624-627.

Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. and Ochman, H., 2005. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences*, 102(47), pp.16919-16926.

Moran, N.A., Dunbar, H.E. and Wilcox, J.L., 2005. Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola. Journal of bacteriology*, 187(12), pp.4229-4237.

Mostovoy, Y., Levy-Sakin, M., Lam, J., Lam, E.T., Hastie, A.R., Marks, P., Lee, J., Chu, C., Lin, C., Džakula, Ž. and Cao, H., 2016. A hybrid approach for de novo human genome sequence assembly and phasing. *Nature methods*, 13(7), p.587.

Müller, C.B., Williams, I.S. and Hardie, J., 2001. The role of nutrition, crowding and interspecific interactions in the development of winged aphids. *Ecological Entomology*, 26(3), pp.330-340.

Munson, M.A., Baumann, P. and Kinsey, M.G., 1991. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated,

primary endosymbionts of aphids. *International Journal of Systematic and Evolutionary Microbiology,* 41(4), pp.566-568.

Myllymäki, H., Valanne, S. and Rämet, M., 2014. The *Drosophila* imd signaling pathway. *The Journal of Immunology*, 192(8), pp.3455-3462.

Nakabachi, A. and Ishikawa, H., 1999. Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *Journal of Insect Physiology*, 45(1), pp.1-6.

Nakamura, T., Yamada, K.D., Tomii, K. and Katoh, K., 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, 34(14), pp.2490-2492.

Nanoporetech, 2018. *Pinfish (v1)* https://github.com/nanoporetech/pipeline-pinfish-analysis.

Ng, J.C. and Perry, K.L., 2004. Transmission of plant viruses by aphid vectors. *Molecular Plant Pathology*, 5(5), pp.505-511.

Norton, G.A. and Mumford, J.D., 1993. Decision Tools for Pest Management. CAB International, Oxon, UK.

Oerke, E.C., 2006. Crop losses to pests. *The Journal of Agricultural Science*, 144(1), pp.31-43.

Oerke, E.C. and Dehne, H.W., 2004. Safeguarding production—losses in major crops and the role of crop protection. *Crop protection*, 23(4), pp.275-285.

Oliver, K.M., Moran, N.A. and Hunter, M.S., 2005. Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proceedings of the National Academy of Sciences*, 102(36), pp.12795-12800.

Oliver, K.M., Russell, J.A., Moran, N.A. and Hunter, M.S., 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, 100(4), pp.1803-1807.

Oliveros, J.C., 2007-2015. Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html

Oliver, S.G., Winson, M.K., Kell, D.B. and Baganz, F., 1998. Systematic functional analysis of the yeast genome. *Trends in biotechnology*, 16(9), pp.373-378.

Oliver, S.V. and Brooke, B.D., 2016. The role of oxidative stress in the longevity and insecticide resistance phenotype of the major malaria vectors *Anopheles arabiensis* and *Anopheles funestus. PloS one*, 11(3), p.e0151049.

Ollivier, M., Gabaldón, T., Poulain, J., Gavory, F., Leterme, N., Gauthier, J.P., Legeai, F., Tagu, D., Simon, J.C. and Rispe, C., 2012. Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that differ by reproductive mode. *Genome biology and evolution*, 4(2), pp.155-167.

Papura, D., Simon, J.C., Halkett, F., Delmotte, F., Le Gallic, J.F. and Dedryver, C.A., 2003. Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity*, 90(5), p.397.

Parker, W.E., 2005. The direct feeding effects of aphids (*Macrosiphum euphorbiae* and *Myzus persicae*) on the yield of potato crops in the UK. *Aspects of Applied Biology*, (76), pp.167-174.

Pearson, W.R., 2013. An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, 42(1), pp.3-1.

Peccoud, J., Loiseau, V., Cordaux, R. and Gilbert, C., 2017. Massive horizontal transfer of transposable elements in insects. *Proceedings of the National Academy of Sciences*, 114(18), pp.4721-4726.

Pedra, J.H., Brandt, A., Li, H.M., Westerman, R., Romero-Severson, J., Pollack, R.J., Murdock, L.L. and Pittendrigh, B.R., 2003. Transcriptome identification of putative genes involved in protein catabolism and innate immune response in human body louse (*Pediculicidae: Pediculus humanus*). *Insect biochemistry and molecular biology*, 33(11), pp.1135-1143.

Phillippy, A.M., 2017. New advances in sequence assembly. *Genome Research*, 27(5), xi-xiii.

Pirone, T.P. and Harris, K.F., 1977. Nonpersistent transmission of plant viruses by aphids. *Annual Review of Phytopathology*, 15(1), pp.55-73.

Pitino, M. and Hogenhout, S.A., 2013. Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Molecular Plant-Microbe Interactions,* 26(1), pp.130-139.

Pockley, A.G., 2003. Heat shock proteins as regulators of the immune response. *The Lancet*, 362(9382), pp.469-476.

Polin, S., Simon, J.C. and Outreman, Y., 2014. An ecological cost associated with protective symbionts of aphids. *Ecology and evolution*, 4(6), pp.836-840.

Powell, S.J. and Bale, J.S., 2004. Cold shock injury and ecological costs of rapid cold hardening in the grain aphid *Sitobion avenae* (*Hemiptera: Aphididae*). *Journal of insect physiology*, 50(4), pp.277-284.

Price, D.R., Feng, H., Baker, J.D., Bavan, S., Luetje, C.W. and Wilson, A.C., 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences*, 111(1), pp.320-325.

Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), pp.559-575.

Qiu, F. and Ungerer, M.C., 2018. Genomic abundance and transcriptional activity of diverse gypsy and copia long terminal repeat retrotransposons in three wild sunflower species. *BMC plant biology*, 18(1), p.6.

Quick, J., 2018. Ultra-long read sequencing protocol for Oxford Nanopore. *Protocols.io.* https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-oxford-nan-k88czzw?version_warning=no.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), pp.841-842.

Raboudi, F., Chavigny, P., Makni, H., Vanlerberghe, F.M. and Makni, M., 2012. Spatial and temporal genetic variation in Tunisian field populations of *Macrosiphum euphorbiae* (Thomas). *Environmental entomology*, 41(2), pp.420-425.

Raboudi, F., Makni, H. and Makni, M., 2011. Genetic diversity of potato aphid, *Macrosiphum euphorbiae*, populations in Tunisia detected by RAPD. *African Entomology*, 19(1), pp.133-141.

Rahbé, Y., Digilio, M.C., Febvay, G., Guillaud, J., Fanti, P. and Pennacchio, F., 2002. Metabolic and symbiotic interactions in amino acid pools of the pea aphid, *Acyrthosiphon pisum*, parasitized by the braconid *Aphidius ervi*. *Journal of Insect Physiology*, 48(5), pp.507-516.

Rambaut, A (2009) FigTree v1.4.3 . *Github.* https://github.com/rambaut/figtree/releases.

Ranson, H. and Hemingway, J., 2005. Mosquito glutathione transferases. *Methods in enzymology*, 401, pp.226-241.

R Core Team (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* http://www.R-project.org/.

Rinehart, J.P., Li, A., Yocum, G.D., Robich, R.M., Hayward, S.A. and Denlinger, D.L., 2007. Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proceedings of the National Academy of Sciences*, 104(27), pp.11130-11137.

Rinehart, J.P., Yocum, G.D. and Denlinger, D.L., 2000. Developmental upregulation of inducible hsp70 transcripts, but not the cognate form, during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. *Insect biochemistry and molecular biology*, 30(6), pp.515-521.

Roach, M.J., Schmidt, S.A. and Borneman, A.R., 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics*, 19(1), p.460.

Robinson, G.E., Hackett, K.J., Purcell-Miramontes, M., Brown, S.J., Evans, J.D., Goldsmith, M.R., Lawson, D., Okamuro, J., Robertson, H.M. and Schneider, D.J., 2011. Creating a buzz about insect genomes. *Science*, 331(6023), pp.1386-1386.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P., 2011. Integrative genomics viewer. *Nature biotechnology*, 29(1), p.24.

Ross, J., Jiang, H., Kanost, M.R. and Wang, Y., 2003. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene*, 304, pp.117-131.

Roth, S., Stein, D. and Nüsslein-Volhard, C., 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell*, 59(6), pp.1189-1202.

Rubio-Melendez, M.E., Barrios-SanMartin, J., Pina-Castro, F.E., Figueroa, C.C. and Ramirez, C.C., 2019. Asexual reproduction of a few genotypes favored the invasion of the cereal aphid *Rhopalosiphum padi* in Chile. *PeerJ*, 7, pp.e7366-e7366.

Russell, J.A. and Moran, N.A., 2005. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proceedings of the Royal Society B: Biological Sciences*, 273(1586), pp.603-610.

Sampson, C., Boullenger, A., Puerto, G. and Hernandez, P., 2011. Implementing Integrated Pest Management programmes in protected strawberry crops across Europe. *IOBC-WPRS Bulletin*, 70, pp.171-180.

Sandström, J.P., Russell, J.A., White, J.P. and Moran, N.A., 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular ecology*, 10(1), pp.217-228.

Sato, D., Sugimoto, M., Akashi, H., Tomita, M. and Soga, T., 2014. Comparative metabolite profiling of foxglove aphids (*Aulacorthum solani* Kaltenbach) on leaves of resistant and susceptible soybean strains. *Molecular BioSystems*, 10(4), pp.909-915.

Scarborough, C.L., Ferrari, J. and Godfray, H.C.J., 2005. Aphid protected from pathogen by endosymbiont. *Science*, 310(5755), pp.1781-1781.

Schmitz, A., Anselme, C., Ravallec, M., Rebuf, C., Simon, J.C., Gatti, J.L. and Poirié, M., 2012. The cellular immune response of the pea aphid to foreign intrusion and symbiotic challenge. *PLoS One*, 7(7), p.e42114.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A. and Minx, P., 2009. The B73 maize genome: complexity, diversity, and dynamics. *science*, 326(5956), pp.1112-1115.

Scholthof, K.B.G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., Hohn, B., Saunders, K., Candresse, T., Ahlquist, P. and Hemenway, C., 2011. Top 10 plant viruses in molecular plant pathology. *Molecular plant pathology*, 12(9), pp.938-954.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), pp.2068-2069.

Seemann T (2015), snippy: fast bacterial variant calling from NGS reads. *GitHub*. https://github.com/tseemann/snippy

Seo, J.S., Rhie, A., Kim, J., Lee, S., Sohn, M.H., Kim, C.U., Hastie, A., Cao, H., Yun, J.Y., Kim, J. and Kuk, J., 2016. De novo assembly and phasing of a Korean human genome. *Nature*, 538(7624), p.243.

Sequeira, R. and Mackauer, M., 1992. Covariance of adult size and development time in the parasitoid wasp *Aphidius ervi* in relation to the size of its host, *Acyrthosiphon pisum. Evolutionary Ecology*, 6(1), pp.34-44.

Shao, H., Ganesamoorthy, D., Duarte, T., Cao, M.D., Hoggart, C.J. and Coin, L.J., 2018. npInv: accurate detection and genotyping of inversions using long read sub-alignment. *BMC bioinformatics*, 19(1), p.261.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H., 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 407(6800), p.81.

Shukla, E., Thorat, L.J., Nath, B.B. and Gaikwad, S.M., 2014. Insect trehalase: physiological significance and potential applications. *Glycobiology*, 25(4), pp.357-367.

Sidney, L.A., Bueno, V.H.P., Lins Jr, J.C., Sampaio, M.V. and Silva, D.B., 2010. Larval competition between *Aphidius ervi* and *Praon volucre* (*Hymenoptera: Braconidae: Aphidiinae*) in *Macrosiphum euphorbiae* (*Hemiptera: Aphididae*). *Environmental entomology*, 39(5), pp.1500-1505.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), pp.3210-3212.

Simon, J.C., Baumann, S., Sunnucks, P., Hebert, P.D., Pierre, J.S., Le Gallic, J.F. and Dedryver, C.A., 1999. Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, 8(4), pp.531-545.

Simon, J.C., Leterme, N. and Latorre, A., 1999. Molecular markers linked to breeding system differences in segregating and natural populations of the cereal aphid *Rhopalosiphum padi* L. *Molecular Ecology*, 8(6), pp.965-973.

Simon, J.C., Rispe, C. and Sunnucks, P., 2002. Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), pp.34-39.

Simon, J.C., Stoeckel, S. and Tagu, D., 2010. Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, 333(6-7), pp.488-496.

Smit, AFA, Hubley, R & Green, P. RepeatMasker. Open-4.0. 2013-2015 http://www.repeatmasker.org.

Smith, R.F., Apple, J.L. and Bottrell, D.G., 1976. The origins of integrated pest management concepts for agricultural crops. *In Integrated pest management* (pp. 1-16). Springer, Boston, MA.

Soderlund, C., Bomhoff, M. and Nelson, W.M., 2011. SyMAP v3. 4: a turnkey synteny system with application to plant genomes. *Nucleic acids research*, 39(10), pp.e68-e68.

Srigiriraju, L., Semtner, P.J., Anderson, T.D. and Bloomquist, J.R., 2010. Monitoring for MACE resistance in the tobacco-adapted form of the green peach aphid, *Myzus persicae* (Sulzer)(*Hemiptera: Aphididae*) in the eastern United States. *Crop Protection*, 29(2), pp.197-202.

Srinivasan, R. and Alvarez, J.M., 2011. Specialized host utilization of *Macrosiphum euphorbiae* on a nonnative weed host, *Solanum sarrachoides*, and competition with *Myzus persicae*. *Environmental entomology*, 40(2), pp.350-356.

Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24(5), pp.637-644.

Stanke. M., Schöffmann, O., Morgenstern, B. and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62.

Strand, M.R., 2014. Teratocytes and their functions in parasitoids. *Current Opinion in Insect Science*, 6, pp.68-73.

Strathdee, A.T., Howling, G.G. and Bale, J.S., 1995. Cold hardiness of overwintering aphid eggs. *Journal of Insect Physiology*, 41(8), pp.653-657.

Sun, Y., Zhao, J., Sheng, Y., Xiao, Y.F., Zhang, Y.J., Bai, L.X., Tan, Y., Xiao, L.B. and Xu, G.C., 2016. Identification of heat shock cognate protein 70 gene (Alhsc70) of *Apolygus lucorum* and its expression in response to different temperature and pesticide stresses. *Insect science*, 23(1), pp.37-49.

Syberg-Olsen, M. and Husnik, F, (2018) Pseudofinder, *GitHub*. https://github.com/filip-husnik/pseudo-finder/.

Tang, Q.L., Ma, K.S., Hou, Y.M. and Gao, X.W., 2017. Monitoring insecticide resistance and diagnostics of resistance mechanisms in the green peach aphid, *Myzus persicae* (Sulzer) (*Hemiptera: Aphididae*) in China. *Pesticide biochemistry and physiology*, 143, pp.39-47.

Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Croft, L.J. and Gan, H.M., 2018. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *GigaScience*, 7(3), p.gix137.

Teixeira, M.A., Sela, N., Atamian, H.S., Bao, E., Chaudhary, R., MacWilliams, J., He, J., Mantelin, S., Girke, T. and Kaloshian, I., 2018. Sequence analysis of the potato aphid *Macrosiphum euphorbiae* transcriptome identified two new viruses. *PloS one*, 13(3), p.e0193239.

Thorpe, P., Escudero-Martinez, C., Cock, P., Laetsch, D., Eves-van den Akker, S. and Bos, J., 2018. Shared transcriptional control and disparate gain and loss of aphid parasitism genes and loci acquired via horizontal gene transfer. *bioRxiv*, p.246801.

Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V. and Kaufman, T.C., 2018. FlyBase 2.0: the next generation. *Nucleic acids research*, 47(D1), pp.D759-D765.

Tomasetto, F., Tylianakis, J.M., Reale, M., Wratten, S. and Goldson, S.L., 2017. Intensified agriculture favors evolved resistance to biological control. *Proceedings of the National Academy of Sciences*, 114(15), pp.3885-3890.

Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., Matsumoto, S., Simon, J.C. and Fukatsu, T., 2010. Symbiotic bacterium modifies aphid body color. *Science*, 330(6007), pp.1102-1104.

Turley, N.E. and Johnson, M.T., 2015. Ecological effects of aphid abundance, genotypic variation, and contemporary evolution on plants. *Oecologia*, 178(3), pp.747-759.

Valanne, S., Wang, J.H. and Rämet, M., 2011. The *Drosophila* toll signaling pathway. *The Journal of Immunology*, 186(2), pp.649-656.

van der Wilk, F., Dullemans, A.M., Verbeek, M. and van den Heuvel, J.F., 1999. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum*. *Virology*, 262(1), pp.104-113.

Van Emden, H.F. and Harrington, R. eds., 2017. *Aphids as crop pests*. CABI, Wallingford, Oxfordshire, UK.

Van Ham, R.C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jiménez, L., Postigo, M., Silva, F.J. and Tamames, J., 2003. Reductive genome evolution in *Buchnera aphidicola. Proceedings of the National Academy of Sciences*, 100(2), pp.581-586.

Van Zee, J.P., Schlueter, J.A., Schlueter, S., Dixon, P., Sierra, C.A.B. and Hill, C.A., 2016. Paralog analyses reveal gene duplication events and genes under positive selection in *Ixodes scapularis* and other *Ixodid* ticks. *BMC genomics*, 17(1), p.241.

Veal, E.A., Toone, W.M., Jones, N. and Morgan, B.A., 2002. Distinct roles for glutathione S-transferases in the oxidative stress response in *Schizosaccharomyces pombe*. *Journal of Biological Chemistry*, 277(38), pp.35523-35531.

Veen, B.W., 1985. Photosynthesis and assimilate transport in potato with top-roll disorder caused by the aphid *Macrosiphum euphorbiae*. *Annals of applied Biology*, 107(2), pp.319-323.

Viant, M.R., Kurland, I.J., Jones, M.R. and Dunn, W.B., 2017. How close are we to complete annotation of metabolomes?. *Current opinion in chemical biology*, 36, pp.64-69.

Via, S., 1991. Specialized host plant performance of pea aphid clones is not altered by experience. *Ecology*, 72(4), pp.1420-1427.

Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S., 2002. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic acids research*, 30(14), pp.3141-3151.

Vorburger, C., 2004. Cold tolerance in obligate and cyclical parthenogens of the peach-potato aphid, *Myzus persicae*. *Ecological Entomology*. 29, 498-505.

Vorburger, C., Ganesanandamoorthy, P. and Kwiatkowski, M., 2013. Comparing constitutive and induced costs of symbiont-conferred resistance to parasitoids in aphids. *Ecology and evolution*, 3(3), pp.706-713.

Vorburger, C., Gehrer, L. and Rodriguez, P., 2009. A strain of the bacterial symbiont *Regiella insecticola* protects aphids against parasitoids. *Biology letters*, 6(1), pp.109-111.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J. and Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14), pp.2202-2204.

Wang, X., Wang, Q. and Qi, Q., 2015. Identification of riboflavin: revealing different metabolic characteristics between *Escherichia coli* BL21 (DE3) and MG1655. *FEMS microbiology letters*, 362(11), fnv071.

Wang, X.W., Luan, J.B., Li, J.M., Su, Y.L., Xia, J. and Liu, S.S., 2011. Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. *BMC genomics*, 12(1), p.458.

Wang, Y., Hereward, J.P. and Zhang, G., 2015. High Spatial Genetic Structure and Genetic Diversity in Chinese Populations of *Sitobion miscanthi* (*Hemiptera: Aphididae*). *Journal of economic entomology*, 109(1), pp.375-384.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), pp.1189-1191.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. and Zdobnov, E.M., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, 35(3), pp.543-548.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M. and Jaffe, D.B., 2017. Direct determination of diploid genome sequences. *Genome research*, 27(5), pp.757-767.

Wellings, P.W., Leather, S.R. and Dixon, A.F.G., 1980. Seasonal variation in reproductive potential: a programmed feature of aphid life cycles. *The Journal of Animal Ecology*, pp.975-985.

Wenger, J.A., Cassone, B.J., Legeai, F., Johnston, J.S., Bansal, R., Yates, A.D., Coates, B.S., Pavinato, V.A. and Michel, A., 2017. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect biochemistry and molecular biology*, 102917, 1-10.

Weng, Y., Perumal, A., Burd, J.D. and Rudd, J.C., 2010. Biotypic diversity in greenbug (Hemiptera: Aphididae): microsatellite-based regional divergence and host-adapted differentiation. *Journal of economic entomology*, 103(4), pp.1454-1463.

Wertheim, B., Kraaijeveld, A.R., Schuster, E., Blanc, E., Hopkins, M., Pletcher, S.D., Strand, M.R., Partridge, L. and Godfray, H.C.J., 2005. Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome biology*, 6(11), p.R94.

Weyel, W. and Wegener, G., 1996. Adenine nucleotide metabolism during anoxia and postanoxic recovery in insects. *Experientia*, 52(5), pp.474-480.

Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramirez-Gonzalez, R.H., De Oliveira, R., Mayer, K.F., Paux, E. and Choulet, F., 2018. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome biology*, 19(1), p.103.

Wick, R.R., Judd, L.M. and Holt, K.E., 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *bioRxiv*, p.543439.

Willis, G.H. and McDowell, L.L., 1982. Pesticides in agricultural runoff and their effects on downstream water quality. *Environmental Toxicology and Chemistry: An International Journal*, 1(4), pp.267-279.

Wilson, A.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P.F., Macdonald, S.J., Schwartz, J.F. and Thomas, G.H., 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola. Insect molecular biology*, 19, pp.249-258.

Xu, J. and James, R.R., 2012. Temperature stress affects the expression of immune response genes in the alfalfa leafcutting bee, *Megachile rotundata. Insect molecular biology*, 21(2), pp.269-280.

Xu, L., Ma, L., Wang, W., Li, L. and Lu, Z., 2019. Phenoloxidases are required for the pea aphid's defence against bacterial and fungal infection. *Insect molecular biology*, 28(2), pp.176-186.

Yamashita, O., Kadono-Okuda, K., Kuwano, E. and Eto, M., 1987. An imidazole compound as a potent anti-ecdysteroid in an insect. *Agricultural and biological chemistry*, 51(8), pp.2295-2297.

Yang, L., Lin, Z., Fang, Q., Wang, J., Yan, Z., Zou, Z., Song, Q. and Ye, G., 2017. The genomic and transcriptomic analyses of serine proteases and their homologs in an endoparasitoid, *Pteromalus puparum*. *Developmental & Comparative Immunology*, 77, pp.56-68.

Yang, X., Thannhauser, T.W., Burrows, M., Cox-Foster, D., Gildow, F.E. and Gray, S.M., 2008. Coupling genetics and proteomics to identify aphid proteins associated with vector-specific transmission of polerovirus (Luteoviridae). *Journal of virology,* 82(1), pp.291-299.

Yeo, S., Coombe, L., Chu, J., Warren, R.L. and Birol, I., 2017. ARCS: assembly roundup by chromium scaffolding. *BioRxiv*, p.100750.

Yukilevich, R., Maroja, L.S., Nguyen, K., Hussain, S. and Kumaran, P., 2018. Rapid sexual and genomic isolation in sympatric *Drosophila* without reproductive character displacement. *Ecology and evolution*, 8(5), pp.2852-2867.

Zamani, A.A., Talebi, A., Fathipour, Y. and Baniameri, V., 2007. Effect of temperature on life history of *Aphidius colemani* and *Aphidius matricariae* (*Hymenoptera: Braconidae*), two parasitoids of *Aphis gossypii* and *Myzus persicae* (*Homoptera: Aphididae*). *Environmental entomology*, 36(2), pp.263-271.

Zhang, Y., Wang, F., Feng, Q., Wang, H., Tang, T., Huang, D. and Liu, F., 2019. Involvement of trehalose-6-phosphate synthase in innate immunity of *Musca domestica*. *Developmental & Comparative Immunology*, 91, pp.85-92.

Zimin, A.V., Puiu, D., Luo, M.C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J. and Salzberg, S.L., 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome research*, 27(5), pp.787-792.